Frits C.R. Spieksma
Combinatorial Optimization Group
Eindhoven University of Technology
f.c.r.spieksma@tue.nl

Inaugural lecture

Practical combinatorial

In his inaugural lecture, delivered on 21 February 2020, Frits Spieksma illustrates the presence of combinatorial optimization in different parts of today's society. From improving the lives of patients with kidney failure to constructing schedules for sports competitions and from finding a seating arrangement in parliaments to scheduling inland waterway traffic, combinatorial optimization plays an instrumental role in all these fields. He sketches these applications, and shows that there is an increased need for incorporating dynamics, fairness and environmental considerations. Scientific research helps in providing satisfying answers to these challenges.

Prologue

This is a set of stories. Each of these stories intends to illustrate the relevance of *combinatorial optimization* in practice. Each of these stories also aims to illustrate the power of mathematical rigor. One story describes how techniques from the field of combinatorial optimization impact the lives of patients with kidney failure for the better. Another story shows how combinatorial optimization can help us to arrive at an objective seating arrangement for newly-elected members of parliament. Yet another story describes how transportation over inland waterways is affected by locks; here, it is shown that the solutions of mathematical models can help in deciding upon the optimal speeds of freight ships, thereby decreasing (CO_2) emissions. A final story features the perhaps mundane subject of scheduling thousands of soccer matches; again, tools from combinatorial optimization allow efficient schedules to be constructed.

Jointly, these stories illustrate the pervasiveness of combinatorial optimization in today's society. From designing train timetables to scheduling classes at high schools and universities, from organizing conferences to analyzing social networks, our world is filled with discrete decisions. The stories that I chose to include here concern recent topics that I have worked on. Many colleagues have been involved, and it was a pleasure and a privilege working with them. And actually, dear reader, after you have read this text, I not only hope to have convinced you that combinatorial optimization is relevant but also to have shown that there is as much relevance as there is fun!

Introduction

While this text may serve as an appetizer to the field of combinatorial optimization, it is neither an introduction to it nor does it contain proofs or theorems. There are excellent books that achieve these goals; I refer to the sources at the end of this section.

It is good, however, to have a basic understanding of combinatorial optimization (CO). The field of CO can be seen as a collection of problems sharing particular characteristics. Two well-known examples of such problems are the assignment problem and the traveling salesman problem; books have been devoted to each of these problems. Here, I use the assignment problem as our leading example to illustrate what the field of CO is about. Indeed, if ever there is a single problem that signifies the success of CO, it is the assignment problem.

In the assignment problem, one is given two n-sets L and R ($n \in \mathbb{N}$) and, for every

Frits Spieksma

optimization

pair of elements $(\ell, r) \in L \times R$, a cost-coefficient $c_{\ell,r}$ is given. The problem is to select n pairs, each consisting of an element from L and an element from R, such that each element of $L \cup R$ is present in some selected pair while the sum of the cost-coefficients of the selected pairs is minimal. The number of solutions for an assignment problems equals $n!$. This property of the assignment problem, i.e. having a finite number of solutions, is in fact the defining property of combinatorial optimization problems: an optimization problem is a *combinatorial optimization problem* when the number of solutions is finite. That means that finding a best solution, i.e. a solution with minimum cost, may not be a trivial task: indeed, enumerating all solutions, even for moderate values of n and using the fastest processor, can be quite impractical, i.e. may take too much time. We have arrived at the basic challenge associated with each CO problem: amidst a finite yet huge number of solutions, we must find a good solution as fast as possible. Returning to the assignment problem, it is an example of a problem that is efficiently solvable: there exists an algorithm that finds an optimum solution in kn^3 computational steps for some $k \in \mathbb{N}$.

In the next four sections, four stories will unfold. As different as the forthcoming stories are, their common denominator is that (i) the optimization problems behind them possess a defining characteristic: the presence of a finite, yet huge number of solutions; and that (ii) the challenge is to find a good solution as fast as possible.

Sources

The trilogy by Schrijver [46] is a landmark in the field. Other books that give an overview of CO are Cook et al. [11], Nemhauser and Wolsey [36], and Korte and Vygen [26]. Although CO is a relatively young field of science, it has a rich history — an account of that history is given in Schrijver [47].

In particular, there is a fascinating story behind the assignment problem that goes back to works of Carl Gustav Jacob Jacobi (1804–1851). I refer to Kuhn [27] for a description of this history, and to Burkard et al. [10] for a book on the assignment problem. A classic book on the traveling salesman problem is the one edited by Lawler et al. [29]; a more recent one is the book by Applegate et al. [2].

An overview of open problems in the field is given by Woeginger [54].

Story 1: kidney exchange

Kidney failure is a disease that affects one in ten persons worldwide. In the age category 65+, it is estimated that one in five persons suffer from kidney failure, explaining why aging societies have a higher prevalence of kidney failure than other societies. In the Netherlands, over 2000 patients are on a waiting list to receive a kidney. The waiting list maintained by EuroTransplant contains over 10 000 patients waiting for a kidney.

The last stage of this disease is called End Stage Renal Disease (ESRD), and a person with ESRD cannot survive without either dialysis or a kidney transplant. The two main causes of kidney failure are diabetes and high-blood pressure; the increasing prevalence of both diabetes and high-blood pressure (caused by lifestyle habits that are only slowly changing), as well as the absence of artificial kidneys, unfortunately ensures that ESRD is a serious, growing disease not likely to disappear.

As it happens, a healthy individual has two kidneys, and can live a normal life with a single kidney. This is a unique feature of this organ and opens the door for the donation of a kidney by a healthy, living individual to a patient with ESRD. The presence of individuals willing to donate a kidney and the presence of patients in need of one gives rise to an allocation problem; I will now describe how the application of sophisticated techniques from CO allows us to redistribute available kidneys from living donors in order to improve the quality of many patients' lives. As a disclaimer, let me point out that gross simplifications are made in this description; more precise and elaborate descriptions can be found in the sources at the end of this section.

Let us consider a person with ESRD, referred to as the patient. In many cases, a patient has found a person (a family member, or a close friend) willing to donate her or his kidney to the patient, opening up the possibility of significantly improving the patient's quality of life. However, a donor must be compatible with the patient in order for the intended transplant to be possible. Compatibility is determined by blood type and by immunological properties of the patient-donor pair. In case the patient-donor pair is incompatible, they may decide to enter a so-called Kidney Exchange Program (KEP). In fact, they may decide to enter a KEP even if they are compatible. Such a program consists of a pool of patient-donor pairs. The idea is that, when viewed from the entering pair (let's call them pair A), there is a possibility that there is one pair present among the donor-patient pairs currently in the pool (let's call them pair B) whose donor is compatible with the patient of pair A, and whose patient is compatible with the donor of pair A. Then, an exchange is possible: the patient of pair A receives a kidney from the donor of pair B, and the patient of pair B receives a kidney from the donor of pair A. The first actual exchange of this type occurred in 1992 in South Korea; since then, this practice has steadily become more popular.

An exchange of this type is called a 2-cycle. Indeed, when building a graph that has a node for each patient-donor pair in the pool and an arc from one node to another node if the donor of the first node is compatible with the patient of the other node, it becomes clear that a set of disjoint 2-cycles in this graph represents a set of exchanges that leave the patients (who correspond to nodes in the graph) with a kidney (and their donors donating one). In fact, a moment of reflection will convince the reader that there is no reason to restrict one's attention to 2-cycles: any cycle in this graph represents a feasible set of transplants.

In practice, however, there are arguments that favor short cycles. Also, in practice, there are persons who are willing to donate a kidney to any patient (so-called non-directed donors), and instead of cycles, this leads to the feature of chains of interest in the graph. An important difference between a chain and a cycle is temporal: while the exchanges that follow from a cycle need to be carried out simultaneously, the exchanges that follow from a chain offer the opportunity to choose well-suited moments for each individual exchange.

Let us now turn to a formal definition of a basic version of the optimization problem in a KEP. An instance of the problem is defined by a simple, directed graph $G=(V,A)$, and by an integer K (denoting the maximum cycle length). Each vertex in V represents a patient-donor pair; an arc $(i,j)\in A$ represents a possible transplant of a kidney from the donor associated with vertex $i\in V$ to the patient associated with vertex $j\in V$. Let C be the set of all directed cycles of, at most, length K and let w_c be the length of cycle c in G. The optimization problem is now to select a set of vertex-disjoint cycles in G which maximizes the total number of arcs contained in the selected cycles. Clearly, this is a problem in CO as the number of solutions is finite. Solving this problem, and its generalizations (which include non-directed donors, weights and uncertainty) is of paramount relevance, and is being done by several (supra-)national institutions, hospitals and special-purpose organizations.

Here is a mathematical formulation of the problem using binary variables z_c, which equal 1 if and only if cycle $c\in C$ is selected.

$$
\begin{aligned}
\text{Maximize} \quad & \sum_{c\in C} w_c z_c && (1)\\
\text{subject to} \quad & \sum_{c:i\in c} z_c \le 1 \quad \text{for each } i\in V, && (2)\\
& z_c\in\{0,1\} \quad \text{for each } c\in C. && (3)
\end{aligned}
$$

Notice that different formulations exist. The one chosen here is the one that I find most elegant to state; it is not the one most suited for computational purposes.

There are many, many challenges in this field. Apart from ethical and medical considerations, it is a fact that different countries have wildly differing rules for organ donation. For instance, only 2-cycles are allowed in France and some countries allow the inclusion of compatible pairs whereas other countries don't. Such issues are relevant as there is an increasing awareness that allowing exchanges across borders may have a positive impact on the quality of the solutions. And, of course, the set of precise rules that govern the exchanges has a huge impact on the solutions possible.

When it comes to mathematical optimization, at least three challenges exist:

– *Dynamics*. A crucial question is: how often should one try to find a set of transplants, i.e. with what frequency should one solve model (1)? Once every month? Or after a sufficient number of changes to the pool? The answer to this question depends on the frequency with which the pool changes; the phenomenon is referred to as 'thickening the market'. Of course, one argument to wait is that after some time, more changes may improve the quality of the solution. An argument not to wait is that any opportunity to realize an additional transplant should not be missed. A fast-growing amount of literature analyzes this situation under various assumptions — it seems that a greedy approach (i.e. solving whenever the situation has changed) is preferable.
– *Recourse*. As described above, the absence of an arc does indeed mean that there is no compatibility between the donor of one pair and the patient of another pair. However, the presence of an arc does not guarantee compatibility; it basically means that there is a good chance that there is compatibility. An in-depth screening of the two persons involved may reveal that the two are incompatible after all. This opens the door to many questions. One variant associates a probability to each arc, representing the likelihood of compatibility. Then, prior to identifying a set of exchanges, one needs to determine which arcs to screen in depth in order to determine their status. Finding these arcs is an optimization question in itself and when the true status of in-depth

screened arcs is revealed, it will have implications for the final solution, i.e. recourse is needed to arrive at a feasible set of exchanges.

– *Fairness*. As the quality of a match depends on properties of a patient, the absence or presence of certain properties may benefit a particular class of patients. More concretely, it has been observed that patients with blood type O have disadvantages in the allocation of kidneys. Also, ethnic imbalances in the allocation may result. Improving models and methods while taking into account an accepted notion of fairness is a major challenge.

Sources

Facts about the prevalence of ESRD can be found in Heaf [22] and on the following sites: https://www.kidney.org/kidneydisease/global-facts-about-kidney-disease and http://www.healthdata.org.

Statistics about the number of kidney transplants are tracked by the World Health Organization [55], https://www.who.int/transplantation/gkt/statistics/en. Numbers for the Dutch situation are maintained by the Dutch Transplant Foundation, see: http://www.transplantatiestichting.nl/publicaties-en-naslag/nts-jaarverslagen.

Elaborate descriptions of the working of a KEP can be found in Gentry et al. [15] and Anderson et al. [1]. For an overview of kidney exchange programs in Europe, see Biró et al. [8]. Mathematical formulations of the optimization problem are surveyed in Mak-Hau [31], see also Glorie et al. [17] and Dickerson et al. [13].

The computational complexity and approximability of the underlying optimization problem is studied in Biró et al. [6].

The issue of thickening the market is discussed in Ashlagi et al. [3] and Ashlagi et al. [4]. Recourse has received attention in Blum et al. [9] and Smeulders et al. [48]. Fairness issues, as well as the impact of a new kidney allocation system employed by UNOS on racial/ethnic disparities in waiting times, are investigated in Glander et al. [16] and Melanson et al. [33].

This story is based on Smeulders et al. [48].

Story 2: seating members of parliament

A key institute in old and modern democracies is the parliament: a collection of persons that have been elected to represent the people and whose main task is to control legislative power. Let us refer to these elected persons as members of parliament or MPs. Almost always, members of parliament are grouped into different parties and members of the same party share a basic point of view on how society should be organized. To exercise their right and duty of controlling power, there is a physical location where the members of parliament meet, discuss, vote, gossip and do all the things that members of parliament are supposed to do. I assume that each MP receives a particular seat in the parliament; this is actually common practice in many parliaments (but not in all: notable exceptions are the House of Representatives in the USA and the House of Commons in the UK. In these parliaments, there is a first-come-first-serve policy for every meeting). The question is how to distribute MPs over the seats in the parliament so as to optimize the functioning of the parliament; let us refer to a distribution of MPs over seats as a *seat-allocation*.

While at first sight the question of who sits where may appear an innocent one, there have been intense debates and rows about this matter in recent years. In fact, it has been argued that as voters become more and more polarized, the views of their chosen representatives seem to diverge more and more. In a number of cases, the discussions have indeed already started with the particular seating that the parliament opts for in a newly-elected parliament. In other words, finding a seat-allocation may be a source of intense debates. In such a belligerent atmosphere, a model-based proposal for the seating of the parliament has a huge advantage over human-made proposals: neutrality.

Another interesting observation which underlines how seat-allocations are far from innocent is the following. The voting behavior of an MP is influenced by the voting behavior of the MP's physical neighbors. Indeed, in Iceland (the country with the oldest parliament functioning today), seats for MPs are allocated randomly, independent of party affiliation. This has created the opportunity to statistically test whether neighbors of an MP have an impact on the MP's voting behavior, and it turns out that the answer is affirmative: not only voting behavior, but even the choice of words is influenced by one's neighbors (see sources at the end of this section).

Let us list a number of properties present in almost all seat-allocations in parliaments all over the world:

– MPs from the same party are seated in clusters. This reflects how communication between MPs of the same party should be an important factor when determining a seat-allocation. More concretely, members of the same party should be seated in each other's vicinity; this allows them to pass information and notes quickly and discreetly. Thus, to facilitate intra-party communication, neighboring seats (to be defined later) are as much as possible allocated to MPs of the same party.
– Not all seats are equally important. Seats in the front have more visibility and allow direct access to the debating spot. Usually, large parties (i.e. parties with more MPs than others) occupy one or more front seats. Seats in the back may be considered less important, while seats with access to a corridor may be preferred over seats between other seats. Note that there are parliaments in which each seat has access to a corridor.
– Many democracies have parties that are labeled somewhere on a 'left-right' spectrum. This left-right positioning is often reflected in the allocation of seats. Phrased more generally, MPs from different parties that are considered to hold similar views are allocated to neighboring positions in the parliament.
– Historical habits: particular seats that are allocated to a specific MP of some party for historical or traditional reasons.

In some parliaments, clear rules exist that yield a seat-allocation. For instance, in the US Senate, senators are ordered by seniority and, starting with the oldest senator, each senator gets to choose a seat with Republicans on the right and Democrats on the left side of the chamber; in Iceland, as mentioned above, a draw determines which MPs sit where. Most parliaments, however, do not have an explicit rule to arrive at a seat-allocation.

I now propose a mathematical model to find a seat-allocation. In doing so, let us primarily focus on the amount of intra-party communication that is possible. However, the model is flexible and can take all kinds of parliament-specific wishes and constraints into account. Thus, the objective here is to find a seat-allocation with maximum communication. A way to precisely model communication between different seats in a parliament is to define the concept of neighboring seats. Let us call two distinct seats *neighboring* when direct contact (in the form of whispering or passing notes) between two MPs in these

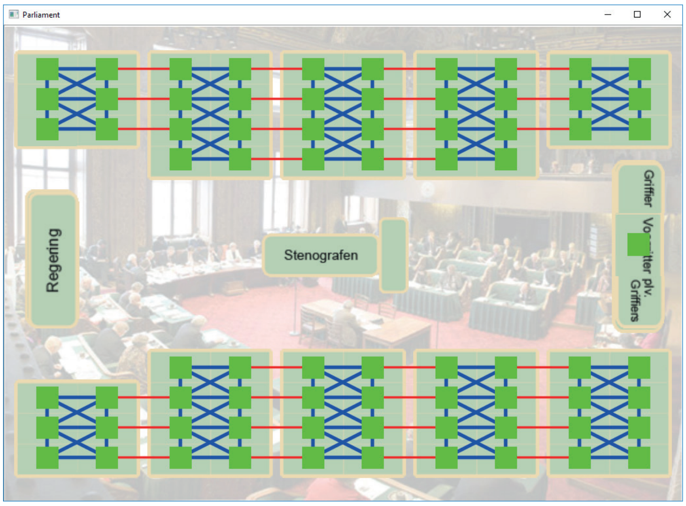

Figure 1 Eerste Kamer, where connections in blue (red) indicate easy (fair) communication.

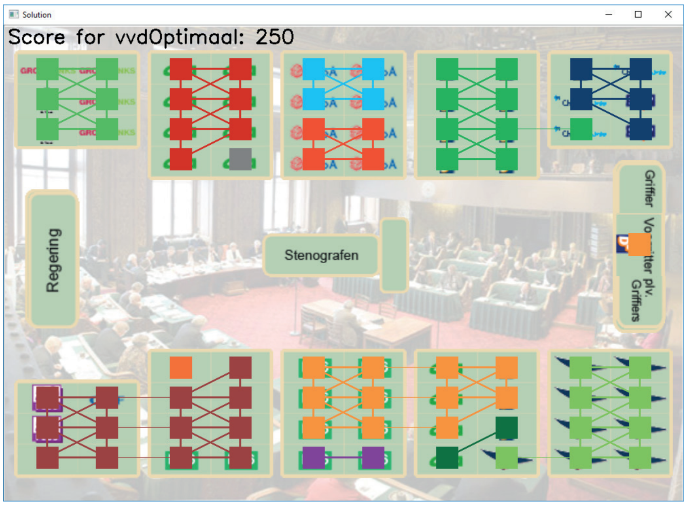

Figure 2 Optimum seat-allocation of the Eerste Kamer.

seats is possible. In practice, this means that two seats that are positioned next to each other on the same row or that are positioned such that one seat is directly behind another seat are called neighboring. Of course, deciding whether two seats are neighboring can be a debatable choice in some cases.

[Again, the reader may skip the next paragraph without consequences for readability]. Let P stand for the set of parties and let S denote the set of seats in the parliament. Party $p \in P$ must receive s_p seats in the parliament. Let us use N to denote the set of pairs of seats that are neighbors, i.e.

$$N = \{(s,t) \in S \times S \mid \text{seats } s \text{ and } t \text{ are distinct neighboring seats}\}.$$

Additionally, let us use binary variables $y_{(s,t),p}$ to indicate whether neighboring seats s and t are assigned to party p and binary variables $x_{s,p}$ to denote whether seat $s \in S$ is assigned to party $p \in P$.

$$\text{Maximize} \quad \sum_{(s,t)\in N}\sum_{p\in P} y_{(s,t),p} \tag{4}$$

$$\text{subject to} \quad \sum_{p\in P} x_{s,p} \leq 1 \quad \forall s \in S, \tag{5}$$

$$\sum_{s\in S} x_{s,p} = s_p \quad \forall p \in P, \tag{6}$$

$$y_{(s,t),p} \leq \min(x_{s,p}, x_{t,p}) \quad \forall s,t \in S \text{ with } (s,t) \in N,\ p \in P, \tag{7}$$

$$x_{s,p} + x_{t,p} - y_{(s,t),p} \leq 1 \quad \forall s,t \in S \text{ with } (s,t) \in N,\ p \in P, \tag{8}$$

$$x_{s,p} \in \{0,1\} \quad \forall s \in S,\ p \in P. \tag{9}$$

This model is an integer programming model that maximizes a particular objective function representing intra-party communication. Objective (4) attempts to allocate pairs of neighboring seats to members of the same party. Let me point out that one may hypothetically wish to achieve an opposite objective in order to counterbalance polarization. Then, one might want to minimize the given objective function (see the discussion above). Furthermore, constraints (5)–(9) ensure that a feasible seat-allocation is found. If desired, many other constraints can be added. Also, one may use more fine-grained ways to define the concept of neighboring seats.

A crucial advantage of a model-based proposal for a seat-allocation is its neutrality; once there is agreement on the basic principles that should guide the seat-allocation, the outcome of the corresponding model should be perceived as unbiased. Admittedly, as far as I am aware, no existing parliament uses optimization techniques to find a seat-allocation; however, the tools are there!

Sources

Seat-allocations have been contested, see an article in the *NRC* [53] or the 'seat wars' described by the BBC [5]. The impact of neighboring MPs on the voting behavior of Icelandic MPs has been described in Saia [45], see also Harmon et al. [21].

Let us illustrate the above by considering a specific parliament: on 20 March 2019 there were provincial elections in the Netherlands. As a consequence, 75 fresh members of the so-called Eerste Kamer were chosen on 27 May 2019; they had their first meeting in mid-June. Here, the following question will be answered: who will sit where? In other words, when maximizing intra-party communication, which member of parliament (MP) will receive which chair?

This question is approached as follows. First, let us determine which pairs of seats in the parliament facilitate *easy* communication (represented by blue links in Figure 1) and which pairs of seats facilitate *fair* communication (represented by red links in Figure 1). When two seats facilitating easy (fair) communication are assigned to members of the same party, 2 (1) points are gained. Of course, the idea is that seat-allocations with more points are better (in terms of communication) than seat-allocations with fewer points. Our objective is to find a seat-allocation with a maximum number of points.

Given the results of the elections for the Eerste Kamer, Figure 2 displays an optimum seat-allocation. This seat-allocation was found by solving an integer program; optimality of the solution was proved using combinatorial arguments. Additional conditions or preferences can easily be incorporated into this integer program, see Tuin [52].

It should be clear that this idea can be applied to any parliament where MPs have a fixed seat. It is interesting to observe that the particular physical lay-out of the seats in the parliament matters for this objective; most parliaments have a (half-)circular shape in which sets of seats are separated by corridors. A very nice overview of the shape and other properties of existing parliaments around the world is found in [56].

This story is based on Spieksma [49].

Story 3: transportation via inland waterways

Since ancient times, rivers have been used as a way of transporting goods and people. History lessons from high school describe how a tribe called the Batavi, while moving along the Rhine, populated a delta around 40 BC now known as the Netherlands. Much older examples exist: documents from before 2500 BC show that the river Nile was routinely used for transporting goods.

And in fact, the Rhine, the Nile and many other rivers and canals are still being used today as means to bring goods and persons from their origin to their destination. Of course, the emergence of motorized transport like cars, trucks, trains and aviation has allowed for much faster and (seemingly) more efficient ways of transporting goods than transport over inland waterways (we will use the phrase 'inland waterways' to denote the set of rivers and canals over which transportation is possible). Yet in recent decades, some advantages of transportation over inland waterways have become more pronounced. Here are some of these advantages (see the end of this section for sources providing more information).

- *Predictability*. While transportation by truck may be hampered by unexpected traffic jams, causing large variations in travel times, inland waterway transportation is extremely reliable. In contrast to transportation over roads where congestion has become the norm, the capacity of rivers and canals has not been reached by far, and there is potential for a significant increase in additional traffic.
- *Safety*. The number of fatal accidents that arise in inland waterway transportation is dwarfed by the deaths arising from road transportation, train transportation and even transportation by air. Transportation over inland waterways is simply the safest mode of transport.
- *Sustainability*. Emissions caused by freight ships are quite low compared to other modes of transportation. The average load capacity of a freight ship equals around 1500 tons; this is equivalent to 60 trucks. As a consequence, the energy consumption of transport over water is less than 20% of that of road transport. It is clear that transport over inland waterways can play a prominent role in the attempts to mitigate the effects of climate change.

These advantages explain the rising amount of transport over inland waterways; this also induces a greater pressure on the associated infrastructure, notably locks. On many inland waterways, locks are required to ensure a suitable water level for navigation and to give ships the opportunity to overcome the difference in water level. In addition, locks can be used to bypass obstacles such as waterfalls or dams, and may serve as protection against floods; the latter type of lock is often found in harbors.

The presence of locks is usually taken for granted. Only when something is amiss does the (economic) relevance of locks becomes clear. I will give two examples:

- On 3 January 2012 a door of a lock in the Twente Canal (the Netherlands) fell down unexpectedly due to metal fatigue. This resulted in economic damage estimated in millions of euro's per week.
- Delayed repairs to the locks in the Mississippi caused an unexpected closure of Lock 52 in the Mississippi river in September 2017, leading to 475 ships being unable to pass. The total costs were estimated at 640 million dollars.

Figure 3 The lock at Ternaaien.

However, even under normal circumstances, locks act as bottlenecks when the waterway traffic density is high and may induce waiting time for ships that pass through these canals and waterways. Efficient lock schedules can contribute to the attractiveness of waterway transport and help to maximize the impact of expensive infrastructural investments.

An optimization problem thus arises when a set of ships has to pass a series of locks along a waterway. A precise description of the corresponding situation is as follows.

Consider a set of $\mathcal{L}=\{1,2,\ldots,L\}$ locks indexed consecutively along a canal or a river. These locks partition the waterway into $L-1$ *segments*. Each lock $\ell\in\mathcal{L}$ has a capacity C_ℓ, indicating how many ships the lock can hold, and has a lockage time P_ℓ, denoting the time that lock ℓ needs to transfer ships through the lock. The distance between locks ℓ and $\ell+1$ is denoted by R_ℓ for each $\ell\in\mathcal{L}\setminus\{L\}$.

Ships arrive at lock 1 and need to pass all locks, with their final lock being lock L; let us use $\mathcal{S}$ to denote the set of ships. Each ship $s\in\mathcal{S}$ arrives at lock 1 at arrival time A_s and must have passed through lock L by the deadline D_s. There is a minimum speed (V_s^{min}) and a maximum speed (V_s^{max}) for each ship $s\in\mathcal{S}$. A ship can pick an individual speed on each segment of the waterway; within a segment, a constant speed of each ship is assumed.

The emissions of a ship are a function of its speed: the faster the ship travels, the more emissions it produces. Literature suggests this relation is cubic; here, I simply assume that there is a given function $E_s(v)$ that describes the emission in tons per kilometer for ship $s\in\mathcal{S}$ that travels with speed v. The goal is to minimize total emissions by selecting a speed for each individual ship while identifying feasible lockages and ensuring that each ship meets its deadline. Time is discretized; let us use $\mathcal{T}$ to denote the set of times at which lockages can start.

Let us use the following variables. Let $v_{s,\ell}$ denote the speed of ship $s \in \mathcal{S}$ at segment $\ell \in \mathcal{L} \backslash \{L\}$. Further, let $x_{s,\ell,t}$ denote a binary variable that equals 1 if lock ℓ starts a lockage containing ship s at time t.

Minimize

$$\sum_{\ell \in \mathcal{L} \backslash \{L\}} R_\ell \sum_{s \in \mathcal{S}} E_s(v_{s,\ell}) \tag{10}$$

subject to

$$\sum_{t \in \mathcal{T}} t x_{s,1,t} \geq A_s \qquad \forall s \in \mathcal{S}, \tag{11}$$

$$\sum_{t \in \mathcal{T}} (t + P_L) x_{s,L,t} \leq D_s \quad \forall s \in \mathcal{S}, \tag{12}$$

$$\sum_{t \in \mathcal{T}} x_{s,\ell,t} = 1 \qquad \forall s \in \mathcal{S}, \ell \in \mathcal{L}, \tag{13}$$

$$\sum_{t \in \mathcal{T}} t x_{s,\ell+1,t} - \sum_{t \in \mathcal{T}} t x_{s,\ell,t} \geq P_\ell + \frac{R_\ell}{v_{s,\ell}} \qquad \forall s \in \mathcal{S}, \ell \in \mathcal{L} \backslash \{L\}, \tag{14}$$

$$x_{s_1,\ell,t_1} + x_{s_2,\ell,t_2} \leq 1 \quad \forall s_1, s_2 \in \mathcal{S}, \ell \in \mathcal{L}, \ \forall t_1 \neq t_2 \text{ with } |t_2 - t_1| \leq 2P_\ell - 1, \tag{15}$$

$$\sum_{s \in \mathcal{S}} x_{s,\ell,t} \leq C_\ell \qquad \forall \ell \in \mathcal{L}, t \in \mathcal{T}, \tag{16}$$

$$V_s^{\min} \leq v_{s,\ell} \leq V_s^{\max} \quad \forall s \in \mathcal{S}, \ell \in \mathcal{L} \backslash \{L\}, \tag{17}$$

$$x_{s,\ell,t} \in \{0,1\} \qquad \forall s \in \mathcal{S}, \ell \in \mathcal{L}, t \in \mathcal{T}. \tag{18}$$

Notice that the model is nonlinear: both the objective function (10) and the constraints (14) are nonlinear.

Observe that the above model can be generalized to include all kinds of practical situations: travel in two directions, the possibility of not allowing ships to overtake one another, locks having multiple chambers. In addition, other objectives, such as flow time, can be modelled.

Sources
The potential of inland waterways for environmentally friendly transport has been addressed in reports of the European Commission (see [14]) and in a report on the New York State Canal System (see Goodban Belt LLC [18]) for the state of New York.

Data concerning the economic damage of lock failures can be found in *The Economist* [38].

Mathematical optimization techniques applied to the scheduling of a single lock can be found in Passchyn et al. [40]; the complexity of scheduling a sequence of locks is addressed in Passchyn and Spieksma [42]. In practice, optimization techniques have been applied to the Welland Canal (Petersen and Taylor [43]), the Kiel Canal (Lübbecke et al. [30], Meisel and Fagerholt [32]) and to the Mississippi (Nauss [35]); see Passchyn [39] for more examples.

This story is based on Passchyn et al. [41].

	Rounds													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
h_1	H	A	H	A	H	A	H	A	H	A	H	A	H	A
h_2	A	H	H	A	H	A	H	H	A	A	H	A	H	A
h_3	A	H	A	A	H	A	H	H	A	H	H	A	H	A
h_4	A	H	A	H	A	A	H	H	A	H	A	H	H	A
h_5	A	H	A	H	A	H	A	H	A	H	A	H	A	H
h_6	H	A	A	H	A	H	A	A	H	H	A	H	A	H
h_7	H	A	H	H	A	H	A	A	H	A	A	H	A	H
h_8	H	A	H	A	H	H	A	A	H	A	H	A	A	H

Table 1 A HAP-set for a league consisting of eight teams.

Rounds						
1	2	3	4	5	6	7
1 vs 2	3 vs 1	1 vs 4	5 vs 1	1 vs 6	7 vs 1	1 vs 8
8 vs 3	5 vs 7	2 vs 3	6 vs 8	2 vs 5	5 vs 3	2 vs 7
7 vs 4	2 vs 6	7 vs 6	4 vs 2	3 vs 4	8 vs 2	3 vs 6
6 vs 5	4 vs 8	8 vs 5	7 vs 3	8 vs 7	6 vs 4	4 vs 5

Table 2 The first seven rounds of a schedule compatible with the HAP-set from Figure 4 where team i has been assigned to h_i, $i = 1, \ldots, 8$.

Story 4: scheduling multiple leagues

Driving around the country on a Saturday or Sunday morning, one is amazed by the amount of activity on sport fields. Soccer, hockey, korfball… thousands of youngsters compete against one another in a seemingly very organized way: referees are present, coaches are present, the teams are present, and everybody seems to fill the right role at the right time. To give an impression of the numbers involved: in the Netherlands alone, there are approximately 500 000 male youngsters and 100 000 female youngsters associated with the Royal Dutch Football Association (KNVB); corresponding numbers for the Royal Dutch Hockey Association (KNHB) are 105 000 female youngsters and 45 000 male youngsters. A large number of these youngsters find their way to the sporting pitch every weekend.

This observation leads to the following question: how does every individual know where to go? How are all these matches chosen on a particular weekend morning? How does one ensure that there aren't six matches scheduled for teams of a club that only has five pitches available?

Clearly, some amount of planning is needed to realize a satisfying competition. A satisfying competition is one that features matches between teams of similar strengths, has limited travel times and ends with a champion and relegation. In this story, a problem is identified that focuses on the capacity of a club, measured by the number of matches that can take place simultaneously at the club's venue.

As testified above, the challenge here is to deal with the dimensions of the problem: thousands of teams belonging to hundreds of clubs need feasible schedules that allow the club's capacities to be satisfied. In order to cope with this huge number of matches, a league organizer typically uses the following approach (see sources at the end of this section). First, a league format and a league size are chosen. Usually, a so-called Double Round Robin format is preferred (meaning that each pair of teams meet twice), with a league size consisting of 6, 8, 10 or 12 teams. Indeed, leagues of even sizes make sense, as otherwise there is a team in each weekend that cannot play. Second, all teams are clustered into leagues of the chosen size. Although it is true that the total number of teams need not be an exact multiple of the league size, an even league size allows the vast majority of the teams to play each round. Furthermore, it is common practice to (i) use a kind of geographical clustering in order to avoid excessive travel distances, (ii) ensure that teams of the same strength/age category are in a same league, and (iii) to avoid teams of the same club being present in the same league. Thirdly, and most crucially for this story, the league organizer assigns teams to so-called Home-Away Patterns (called a HAP-set, see Table 1 for an illustration). From this assignment, a feasible schedule follows. As an illustration of the latter procedure, consider the HAP-set depicted in Table 1; it reflects a specific HAP-set for a league consisting of eight teams. Although a priori, different schedules (or none) might be compatible with the given HAP-set, Table 2 simply gives one such schedule which is compatible with the HAP-set from Table 1.

To summarize this discussion: in order to efficiently find schedules for many leagues, the league organizer (assuming a given partition of teams into leagues, and the presence of a HAP-set) uses a two-step approach:

- Step 1: assign teams to HAPs.
- Step 2: next, use a table such as the one depicted in Table 2 to find the actual schedule.

Here, the focus is on Step 1. Since an assignment of teams to HAPs dictates when each team plays home, a solution to Step 1 specifies how many matches are played at the club's venue in each round for each club. This is important as the capacity of a club in terms of the number of matches it can host in a round is typically bounded. In fact, it is assumed that a capacity is given for each club in each round; in practice, this number follows from the number of available pitches, the set of possible starting times and the availability of material and referees. The goal is to find, for each league, an assignment of teams to HAPs which minimizes the total capacity violation across the clubs.

We describe the problem formally. A set T of teams ($n=|T|$), a set L of leagues ($m=|L|$) and a set C of clubs are given. Also given are two partitions of the set T; one partition $\{\bar{T}_1,\ldots,\bar{T}_m\}$ of the set T indicates which teams belong to which league; notice that, for each $\ell\in L$, $|\bar{T}_\ell|=k$, with k even since each league consists of the same even number of teams. Another partition of the set T is $\{\hat{T}_1,\ldots,\hat{T}_{|C|}\}$, describing which teams belong to which club. In addition, k HAPs are given, each of length $2(k-1)$, which jointly form a feasible, complementary HAP-set denoted by $\mathcal{H}$. The set of rounds is $\{1,2,\ldots,2(k-1)\}$ and is denoted by R. Finally, each club $c\in C$ has a given capacity $\delta_{c,r}$, which corresponds to the number of matches club c can host in round r.

Capacity violations happen whenever the number of teams of a club that play home in some round exceeds the club's capacity in that round. The violation of a club in a round is measured by a scalar value that is either zero (if there is no violation) or equal to the number of teams that play home in that round minus the club's capacity (if there is a violation).

The multi-league sports scheduling problem (MSP) is now to find an assignment of teams to HAPs, such that the total capacity violation (i.e. the summation of violations over all clubs and all rounds) is minimized.

Let us introduce binary variables $x_{t,h}$, which equal one if team $t\in T$ is assigned to HAP $h\in\mathcal{H}$ and zero otherwise, and auxiliary variables $z_{c,r}$ that represent the amount of violations for club $c\in C$ in round $r\in R$. An assignment is feasible if and only if the teams in each league are assigned to different HAPs. Given the set of HAPs and the set of rounds, one can compute (in a pre-processing step) parameters $U_{h,r}$, which equal one if the team assigned to HAP $h\in\mathcal{H}$ plays home in round $r\in R$ and zero otherwise. The following mixed integer program formulates MSP.

$$\text{Minimize} \sum_{c\in C}\sum_{r\in R} z_{c,r} \tag{19}$$

subject to

$$\sum_{t\in\bar{T}_\ell} x_{t,h}=1 \quad \forall \ell\in L, h\in\mathcal{H}, \tag{20}$$

$$\sum_{h\in\mathcal{H}} x_{t,h}=1 \quad \forall t\in\bar{T}_\ell, \ell\in L, \tag{21}$$

$$z_{c,r}\geq \max\left(\sum_{t\in\hat{T}_c}\sum_{h\in\mathcal{H}} x_{t,h}U_{h,r}-\delta_{c,r},0\right) \quad \forall c\in C, r\in R, \tag{22}$$

$$x_{t,h}\in\{0,1\} \quad \forall t\in T, h\in\mathcal{H}. \tag{23}$$

Various extensions of this model can be relevant as well. The presence of HAP-sets that are league-specific, the existence of leagues of different sizes or existing constraints between specific teams of a club are all examples of meaningful extensions.

Sources

Information about membership numbers can be found in KNVB [24] and KNHB [23]. Toffolo et al. [50] discuss the problem of grouping teams into leagues. The issue of deciding whether a schedule exists for a given HAP-set is a well-researched topic, see Miyashiro et al. [34] and Goossens and Spieksma [20].

Optimization techniques that schedule a single league have been used abundantly in practice, most notably in professional soccer (Goossens and Spieksma [20]) and baseball (Trick et al. [51]), see Knust [25] for an overview of this literature.

The HAP-set portrayed in Table 1 is an example of a so-called flexible HAP-set, see Lambers et al. [28].

This story is based on Davari et al. [12].

Epilogue

A field of science which is relevant is neither immune nor oblivious to developments in practice. The last decade has seen a tremendous increase in the availability of data and in the power to infer structures and insights from these data. In fact, one could argue that the continuous dissemination of rational/scientific solutions to address needs in society has sparked this growth in data availability.

In any case, these detailed data offer the potential to compute solutions that incorporate, more so than before, properties from at least two themes: (i) dynamics and (ii) fairness. Receiving routing advice on your daily commute in order to circumvent a traffic jam caused by an accident that happened ten minutes ago, for instance, is an example of the inclusion of dynamic properties. Likewise, detailed knowledge about the properties of the existing allocation procedures in kidney exchange can help to establish what fairness means in this context and offers a way to increase fairness in kidney allocation.

In the four stories above, I have shown how practical combinatorial optimization can be. However, this should not be taken as an argument against theory. To achieve relevance, a deep understanding of the theoretical properties of combinatorial optimization problems is needed. Indeed, the perceived quality of solutions used in practice will depend on our ability to incorporate new aspects such as dynamics and fairness. To be able to do so, we require new insights and new theory on what it means to be fair and/or dynamic. Models will be fine-tuned and new techniques (such as machine learning) will be investigated and better understood. I am looking forward to working with the members of the Combinatorial Optimization Group to further contribute to these developments!

References

1 R. Anderson, I. Ashlagi, D. Gamarnik, M. Rees, A. Roth, T. Sönmez and M. Ünver, Kidney exchange and the alliance for paired donation: Operations research changes the way kidneys are transplanted, *Interfaces* 45 (2015), 26–42.

2 D. Applegate, R. Bixby, V. Chvátal and W. Cook, *The Traveling Salesman Problem. A Computational Study*, Princeton University Press, 2006.

3 I. Ashlagi, A. Bingaman, M. Burq, V. Manshadi, D. Gamarnik, C. Murphey, A. Roth, M. Melcher and M. Rees, Effect of matchrun frequencies on the number of transplants and waiting times in kidney exchange, *American Journal of Transplantation* 18 (2018), 1177–1186.

4 I. Ashlagi, M. Burq, P. Jailleta and V. Manshadi, On matching and thickness in heterogeneous dynamic markets, *Operations Research* 67 (2019), 927–949.

5 BBC, https://www.bbc.com/news/uk-scotland-scotland-politics-32802374 (2015), accessed at 2 January 2020.

6 P. Biró, D. Manlove and R. Rizzi, Maximum weight cycle packing in directed graphs, with application to kidney exchange programs, *Discrete Mathematics, Algorithms and Applications* 1 (2009), 499–517.

7 P. Biró, B. Haase-Kromwijk, T. Andersson, E. Ásgeirsson, T. Baltesová T, I. Boletis, C. Bolotinha, G. Bond, G. Böhmig, L. Burnapp, K. Cechlárová, P. Di Ciaccio, J. Fronek, K. Hadaya, A. Hemke, C. Jacquelinet, R. Johnson, R. Kieszek, D. Kuypers, R. Leishman, M. Macher, D. Manlove, G. Menoudakou, M. Salonen M, B. Smeulders, V. Sparacino, F. Spieksma, M. de la Oliva Valentïn Muñoz, N. Wilson and J. van de Klundert, Building kidney exchange programmes in Europe — an overview of exchange practice and activities, *Transplantation* 103 (2019), 1514–1522.

8 P. Biró, J. van de Klundert, D. Manlove, W. Pettersson, T. Andersson, L. Burnapp, P. Chromy, P. Delgado, P. Dworczak, B. Haase, A. Hemke, R. Johnson, X. Klimentova, D. Kuypers, A. Costa, B. Smeulders, F. Spieksma, M. Valentín and A. Viana, Modelling and optimisation in European kidney exchange programmes, to appear in the *European Journal of Operational Research* (2019).

9 A. Blum, J. Dickerson, N. Haghtalab, A. Procaccia, T. Sandholm and A. Sharma, Ignorance is almost bliss: Near-optimal stochastic matching with few queries, *Proceedings of the Sixteenth ACM Conference on Economics and Computation* (2015), 325–342.

10 R. Burkard, M. Dell'Amico and S. Martello, *Assignment Problems*, SIAM, 2009.

11 W. Cook, W. Cunningham, W. Pulleyblank and A. Schrijver, *Combinatorial Optimization*, Wiley, 1998.

12 M. Davari, D. Goossens, J. Beliën, R. Lambers and F.C.R. Spieksma, Multi-League Sport Scheduling, or how to schedule thousands of matches (2019), submitted for publication.

13 J. Dickerson, D. Manlove, B. Plaut, T. Sandholm and J. Trimble, Position-indexed formulations for kidney exchange, *Proceedings of the 2016 ACM Conference on Economics and Computation* (2016), 25–42.

14 European Commission, Promotion of inland waterway transport, http://ec.europa.eu/transport/inland/promotion/promotion-en.htm (2015).

15 S. Gentry, R. Montgomery and D. Segev, Kidney paired donation: Fundamentals, limitations, and expansions, *American Journal of Kidney Diseases* 57 (2011), 144–151.

16 P. Glander, K. Budde, D. Schmidt, T. Florian Fuller, M. Giessing, H. Neumayer and L. Liefeldt, The 'blood group O problem' in kidney transplantation — time to change?, *Nephrology Dialysis Transplantation* 25 (2010), 1998–2004.

17 K. Glorie, B. Haase-Kromwijk, J. van de Klundert, A. Wagelmans and W. Weimar, Allocation and matching in kidney exchange programs, *Transplant International* 27 (2014), 333–343.

18 Goodban Belt, New York state canal system — modern freightway, Technical Report, New York State, 2010.

19 D. Goossens and F.C.R. Spieksma, Breaks, cuts, and patterns, *Operations Research Letters* 39 (2011), 428–432.

20 D. Goossens and F.C.R. Spieksma, Soccer schedules in Europe: an overview, *Journal of Scheduling* 15 (2012), 641–651.

21 N. Harmon, R. Fisman and E. Kamenica, Peer effects in legislative voting, Working paper, 2018.

22 J. Heaf, Current trends in European renal epidemiology, *Clinical Kidney Journal* 10 (2017), 149–153.

23 KNHB, Cijfers en plannen, https://www.knhb.nl/over-knhb/cijfers-en-plannen (2018), in Dutch.

24 KNVB, KNVB in cijfers 2017/'18, https://www.knvb.h5mag.com/jaarverslag-2018-18/de-knvb-in-cijfers (2018), in Dutch.

25 S. Knust, Classification of literature on sports scheduling, http://www2.informatik.uni-osnabrueck.de/knust/sportssched/sportlit-class (2019).

26 B. Korte and J. Vygen, *Combinatorial Optimization: Theory and Algorithms*, Springer, 2006.

27 H.W. Kuhn, A tale of three eras: The discovery and rediscovery of the Hungarian Method, *European Journal of Operational Research* 219 (2012), 641–651.

28 R. Lambers, D. Goossens and F.C.R. Spieksma, On the flexibility of Home-Away pattern sets (2019), submitted for publication.

29 E. Lawler, J.K. Lenstra, A. Rinnooy Kan and D. Shmoys, eds., *The Traveling Salesman Problem*, Wiley, 1985.

30 E. Lübbecke, M. Lübbecke and R. Möhring, Ship traffic optimization for the Kiel Canal, *Operations Research* 67 (2019), 791–812.

31 V. Mak-Hau, On the kidney exchange problem: cardinality constrained cycle and chain problems on directed graphs: a survey of integer programming approaches, *Journal of Combinatorial Optimization* 33 (2017), 35–59.

32 F. Meisel and K. Fagerholt, Scheduling two-way ship traffic for the Kiel Canal: Model, extensions and a matheuristic, *Computers & Operations Research* 106 (2019), 119–132.

33 T. Melanson, J. Hockenberry, L. Plantinga, M. Basu, S. Pastan, S. Mohan, D. Howard and R. Patzer, New kidney allocation system associated with increased rates of transplants among black and hispanic patients, *Health Affairs* 36 (2017), 1078–1085.

34 R. Miyashiro, H. Iwasaki and T. Matsui, Characterizing feasible pattern sets with a minimum number of breaks, in: *Proceedings of the 4th International Conference on the Practice and Theory of Automated Timetabling (PATAT2002)*, Lecture Notes in Computer Science 2740, 2003, pp. 78–99.

35 R. Nauss, Optimal sequencing in the presence of setup times for tow/barge traffic through a river lock, *European Journal of Operational Research* 187 (2008), 1268–1281.

36 G. Nemhauser and L. Wolsey, *Integer and Combinatorial Optimization*, Wiley, 1988.

37 NTS, Jaarverslag 2018 — Nieuwe kansen omarmen, http://www.transplantatiestichting.nl/publicaties-en-naslag/nts-jaarverslagen (2018), in Dutch.

38 *The Economist*, Take me to the river, October 2017.

39 W. Passchyn, *Scheduling Locks on Inland Waterways*, PhD thesis, KU Leuven, 2016.

40 W. Passchyn, D. Briskorn, S. Coene, J. Hurink, F.C.R. Spieksma and G. Vanden Berghe, The lockmaster's problem, *European Journal of Operational Research* 251 (2016), 432–441.

41 W. Passchyn, D. Briskorn and F.C.R. Spieksma, Mathematical programming models for lock scheduling with an emission objective, *European Journal of Operational Research* 248 (2016), 802–814.

42 W. Passchyn and F.C.R. Spieksma, Scheduling parallel batching machines in a sequence, *Journal of Scheduling* 22 (2019), 335–357.

43 E.R. Petersen and A.J. Taylor, An optimal scheduling system for the Welland Canal, *Transportation Science* 22 (1988), 173–185.

44 R. Rasmussen and M. Trick, Round robin scheduling: A survey, *European Journal of Operational Research* 188 (2008), 617–636.

45 A. Saia, Random interactions in the Chamber: Legislators' behavior and political distance, *Journal of Public Economics* 164 (2018), 225–240.

46 A. Schrijver, *Combinatorial Optimization*, Springer, 2003.

47 A. Schrijver, On the history of combinatorial optimization (till 1960), homepages.cwi.nl/~lex/files/histco.pdf.

48 B. Smeulders, V. Bartier, Y. Crama and F.C.R. Spieksma, Recourse in kidney exchange programs, 2019, submitted.

49 F.C.R. Spieksma, Who sits where in the Senate?, https://www.networkpages.nl/who-sits-where-in-the-senate, *The Network Pages*, (2019).

50 T. Toffolo, J. Christiaens, F.C.R. Spieksma and G. Vanden Berghe, The sport teams grouping problem, *Annals of Operations Research* 275 (2019), 223–243.

51 M.A. Trick, H. Yildiz and T. Yunes, Scheduling major league baseball umpires and the traveling umpire problem, *Interfaces* 42 (2012), 232–244.

52 D. Tuin, *Seating the parliament*, Bachelor Thesis, Eindhoven University of Technology, 2019.

53 P. Wijnen, Stoelendans in nieuwe Tweede Kamer blijft stuiten op bezwaren Denk en Forum, *NRC*, https://www.nrc.nl/nieuws/2017/03/22/stoelendans-in-nieuwe-tweede-kamer-blijft-stuiten-op-bezwaren-denk-en-forum-7516272-a1551485, 2017.

54 G.J. Woeginger, Some easy and some not so easy geometric optimization problems, in: *Proceedings of the 16th International Workshop on Approximation and Online Algorithms (WAOA2018)*, 2018, pp. 1–12.

55 World Health Organization, Global knowledge base on transplantation, https://www.who.int/transplantation/gkt/statistics/en, 2019.

56 XML, *Parliament*, parliamentbook.com/book.