UNIVERSITY OF VERONA

Master’s degree in Data Science

Discrete optimization and decision making

Course project: The Examination Timetabling Problem

2022/2023

Michael Fidanza

Table of Contents

Introduction

This document's aim is to summarize the work carried out for the module "Discrete Optimization and Decision Making" in the context of the Master's degree in Data Science provided by the University of Verona.
The description of a "real life" problem is given with the purpose of finding a suitable Integer Linear programming formulation and, ultimately, solve it through the usage of Gurobi's library.

Examination Timetabling Problem (ETP)

Let us consider a set E of exams, to be scheduled during an examination period at the end of the semester, and a set S of students. Each student is enrolled in a non-empty subset of exams. The examination period is divided into T ordered time-slots.
Given two exams $e1, e2 \in E$, let $n_{e1,e2}$ be the number of students enrolled in both.
Two exams $e1, e2 \in E$ are called conflicting if they have at least one student enrolled in both, i.e., if $n_{e1,e2} > 0$.
Rules and regulations impose that conflicting exams cannot take place in the same time- slot. Moreover, to promote the creation of timetables more sustainable for the students, a penalty is assigned for each pair of conflicting exams scheduled up to a distance of 5 time-slots. More precisely, given two exams $e1, e2 \in E$ scheduled at distance i of time-slots, with $1 \leq i \leq 5$, the relative penalty is $2^{(5-i)} \cdot \frac{n_{e1,e2}}{|S|}$

The ETP aims at assigning exams to time-slots ensuring that:

- Each exam is scheduled exactly once during the examination period.
- Two conflicting exams are not scheduled in same time-slot.
- The total penalty resulting from the created timetable is minimized.

Base ILP formulation

Given

- $\boldsymbol{E} = \{\boldsymbol{e_1}, \boldsymbol{e_2}, \dots, \boldsymbol{e_n}\}$ set of n exams
- $\boldsymbol{S} = \{\boldsymbol{s_1}, \boldsymbol{s_2}, \dots, \boldsymbol{s_m}\}$ set of m students
- $\boldsymbol{T} = \{\boldsymbol{t_1}, \boldsymbol{t_2}, \dots, \ \boldsymbol{t_l}\}$ set of l time-slots

It's easy to compute

- $\boldsymbol{n_{ee'}}$ as the number of students which are enrolled in both exams e and e'

Decision Variables

- $\boldsymbol{x_{et}} = \mathbf{1}$ if exam e is scheduled in time-slot t, $\forall\ e \in E, t \in T$
- $\boldsymbol{y_{ete't'}} = \mathbf{1}$ if exam e is scheduled in time-slot t and exam e' is scheduled in time-slot t', $\forall\ e, e' \in E, n_{ee'} > 0,\ t, t' \in T$

Objective function

$$\min \Big(\sum_{\substack{e,e' \in E \\ e' > e \\ n_{ee'} > 0}} \ \sum_{\substack{t,t' \in T \\ |t-t'| \le 5}} \frac{2^{5-|t-t'|} \cdot \boldsymbol{n_{ee'}} \cdot \boldsymbol{y_{ete't'}}}{\boldsymbol{m}} \Big)$$

Constraints

- Each exam is scheduled exactly once during the examination period.
  - $\sum_{t \in T} \boldsymbol{x_{et}} = \mathbf{1} \qquad \forall e \in E$
- Two conflicting exams can not be scheduled in same time-slot.
  - $\boldsymbol{x_{et}} + \boldsymbol{x_{e't}} \le \mathbf{1} \qquad \forall\ e, e' \in E, e' > e, n_{ee'} > 0, \forall\ t \in T$
- Link variable y to variable x.
  - $\boldsymbol{y_{ete't'}} \ge \boldsymbol{x_{et}} + \boldsymbol{x_{e't'}} - \mathbf{1} \qquad \forall\ e, e' \in E, n_{ee'} > 0, \forall\ t, t' \in T$
  - $\boldsymbol{y_{ete't'}} \le \boldsymbol{x_{et}} \qquad \forall\ e, e' \in E, n_{ee'} > 0, \forall\ t, t' \in T$
  - $\boldsymbol{y_{ete't'}} \le \boldsymbol{x_{e't'}} \qquad \forall\ e, e' \in E, n_{ee'} > 0, \forall\ t, t' \in T$
- Domain constraints
  - $\boldsymbol{x_{et}} \in \{\mathbf{0}, \mathbf{1}\} \qquad \forall\ e, \in E, \forall\ t \in T$
  - $\boldsymbol{y_{ete't'}} \in \{\mathbf{0}, \mathbf{1}\} \qquad \forall\ e, e' \in E, \forall\ t, t' \in T$

Equity measure 1: Maximum average distance between conflicting exams

Description

In this variation of the problem, the objective function is modified to maximize the average distance between conflicting exams in the examination period.
This measure should be more equitable for students who are enrolled in more than one exam, as they should have, on average, more time between those exams.
Any non-specified constraints and variable definitions are kept equal to the base formulation.

Objective Function

Define $\boldsymbol{num\_conflicts}$ as the total number of conflicts between all possible pair of exams.

$$\max \left(\sum_{\substack{e,e' \in E \\ e' > e \\ n_{ee'} > 0}} \sum_{t,t' \in T} \frac{|t - t'| \cdot y_{ete't'}}{num\_conflicts} \right)$$

Equity measure 2: Minimum number of students with back-to-back exams

Description

In this variation of the problem, the objective function is modified to minimize the number of students who have at least two conflicting exams scheduled in consecutive time-slots.
This measure should be equitable for students who enrolled in more than one exam, as the model will try to provide at least one day off between conflicting exams.
Any non-specified constraints and variable definitions are kept equal to the base formulation.

Objective Function

$$\mathbf{min}(\sum_{s \in S} \boldsymbol{z_s})$$

Decision Variables

- $\boldsymbol{z_s} = \mathbf{1}$ if student $\boldsymbol{s}$ is enrolled at least in two exams which are scheduled in consecutive time slots, $\forall\, s \in S$.

Constraints

Define $\boldsymbol{enr_s}$ as the list of exams in which student $\boldsymbol{s}$ is enrolled.

- Link variable $\boldsymbol{z}$ to $\boldsymbol{y}$
  - $\boldsymbol{z_s} \geq \boldsymbol{y_{ete't'}}$ $\qquad \forall s \in S, \forall e, e' \in enr_s, e \neq e', t \in \{1,2, \dots, |T| - 1\}, t' = t + 1$

  This constraint will force the variable $\boldsymbol{z_s}$ to get value 1 if at least two exams in which student $\boldsymbol{s}$ is enrolled, are scheduled in consecutive time slots. Otherwise, it will be forced to 0 by the optimization function.

- Domain constraints
  - $\boldsymbol{z_s} \in \{\mathbf{0}, \mathbf{1}\}$ $\qquad \forall\, s \in S$

Equity measure 3: Maximize minimum distance between any two conflicting exams

Description

In this variation of the problem, the objective function is modified to maximize the distance between any two conflicting exams.
This measure should be equitable for students who enrolled in more than one exam, as the model will provide a certain minimum number of time slots (based on the length of the examination period and on the number of conflicts) between any two conflicting exams.

Objective Function

$$\mathbf{max}\ (\mathbf{w})$$

Decision Variables

- $\boldsymbol{w}$ integer variable representing the minimum distance between any two conflicting exams.

Constraints

- Define $\boldsymbol{w}$ as
  - $\boldsymbol{w} \leq \sum_{\boldsymbol{t},\boldsymbol{t}' \in \boldsymbol{T}} |\boldsymbol{t} - \boldsymbol{t}'| * \boldsymbol{y}_{\boldsymbol{ete}'\boldsymbol{t}'} \qquad \forall e, e' \in E,\ e' > e, n_{ee'} > 0$

  This constraint, together with the objective function, will force variable $\boldsymbol{w}$ to get the minimum value between all the distances among conflicting exams.

- Domain constraints
  - $\boldsymbol{w} \geq \mathbf{0} \qquad \boldsymbol{Integer}$

Equity measure 4: Maximize average distance between conflicting exams adjusted with Mean Absolute Deviation

Description

In this variation of the problem, the objective function is modified to maximize the average distance between conflicting exams trying also to minimize the value of the mean absolute deviation.
This measure should provide an improvement to the equity measure seen at page 6, as it will try to keep the variability of the distances between conflicting exams low.
The mean absolute deviation refers to the average distance between a specific data point and the mean of all the data points.

Objective Function

$$\max \left(w1 * \sum_{\substack{e,e' \in E \\ e' > e \\ n_{ee'} > 0}} \sum_{t,t' \in T} \frac{|t - t'| \cdot y_{ete't'}}{num\_conflicts} - w2 * \sum_{\substack{e,e' \in T \\ e' > e \\ n_{ee'} > 0}} \frac{d_{ee'}}{num\_conflicts}\right)$$

With $w1, w2$ weights to be determined empirically.

Decision Variables

- $d_{ee'}$ $\forall e, e' \in E$ $e' > e, n_{ee'} > 0$, is an integer variable representing the absolute value of the difference between the distance among conflicting exams e and $e'$and the average distance among all conflicting exams.

Constraints

- Define $d_{ee'}$ as
  - $d_{ee'} \geq \sum_{t,t' \in T} |t - t'| * y_{ete't'} - \sum_{\substack{e,e' \in E \\ e' > e \\ n_{ee'} > 0}} \sum_{t,t' \in T} \frac{|t-t'| \cdot y_{ete't'}}{num\_conflicts}$ $\quad \forall e, e' \in E,$ $e' > e,\ n_{ee'} > 0$
  - $d_{ee'} \geq \sum_{\substack{e,e' \in E \\ e' > e \\ n_{ee'} > 0}} \sum_{t,t' \in T} \frac{|t-t'| \cdot y_{ete't'}}{num\_conflicts} - \sum_{t,t' \in T} |t - t'| * y_{ete't'}$ $\quad \forall e, e' \in E,$ $e' > e,\ n_{ee'} > 0$

  These constraints represent the linearization of the absolute value required in computing MAD measure.

- Domain constraints
  - $d_{ee'} \geq 0$ $\quad$ $Integer$ $\quad$ $\forall e, e' \in E, e' > e,\ n_{ee'} > 0$

Additional restriction 1: At most 3 consecutive time slots with conflicting exams

Description

Add a new restriction which limits the number of consecutive time slots with conflicting exams to a maximum of 3.
Here we are interested in looking for conflicts only in consecutive time slots.

Decision Variables

- $\boldsymbol{k_{tt'}} = \mathbf{1} \qquad \forall\, t \in \{1, 2, \dots, |T| - 1\}, t' \in \{t + 1, \dots, |T|\}$
  if at least a pair of conflicting exams is scheduled in time slots $\boldsymbol{t}$ and $\boldsymbol{t'}$.

Constraints

- Link variable $\boldsymbol{k}$ to $\boldsymbol{y}$
  - $\boldsymbol{k_{tt'}} \leq \sum_{\substack{\boldsymbol{e,e' \in E} \\ \boldsymbol{n_{ee'}>0}}} \boldsymbol{y_{ete't'}} \qquad \forall\, t \in \{1, 2, \dots, |T| - 1\},\ t' \in \{t + 1, 2, \dots, |T|\}$
  - $\boldsymbol{k_{tt'}} \geq \dfrac{\sum_{\substack{\boldsymbol{e,e' \in E} \\ \boldsymbol{n_{ee'}>0}}} \boldsymbol{y_{ete't'}}}{\boldsymbol{num_{conflicts}}} \qquad \forall\, t \in \{1, 2, \dots, |T| - 1\},\ t' \in \{t + 1, 2, \dots, |T|\}$

  These constraints force $\boldsymbol{k_{tt'}}$ to have value 0 in case there are no conflicting exams scheduled in time slots $\boldsymbol{t}$ and $\boldsymbol{t'}$, and value 1 if at least a pair of conflicting exams is scheduled in those time slots. $\boldsymbol{num\_conflicts}$ is an upper bound to the possible number of actual conflicts, therefore the division $(\sum_{\substack{\boldsymbol{e,e' \in E} \\ \boldsymbol{n_{ee'}>0}}} \boldsymbol{y_{ete't'}})/\boldsymbol{num\_conflicts}$ gives back a value between 0 (excluded) and 1 whenever at least a pair of conflicting exams is scheduled in $\boldsymbol{t}$ and $\boldsymbol{t'}$.

- Can't have more than 3 consecutive time slots with conflicting exams.
  - $\boldsymbol{k_{t+2,t+3}} \leq \mathbf{2} - (\boldsymbol{k_{t,t+1}} + \boldsymbol{k_{t+1,t+2}}) \qquad \forall\, t \in \{1,2, \dots, |T| - 3\}$
  - $\boldsymbol{k_{t+3,t+4}} \leq \mathbf{2} - (\boldsymbol{k_{t,t+1}} + \boldsymbol{k_{t+1,t+2}}) \qquad \forall\, t \in \{1,2, \dots, |T| - 4\}$

  With these constraints, in case we have 3 consecutive time slots having conflicting exams scheduled (e.g. t, t+1 and t+2), then we force the 4th time slot (t+3) to not have conflicts with his adjacent time slots (t+2 and t+4).

- Domain constraints
  - $\boldsymbol{k_{tt'}} \in \{\mathbf{0}, \mathbf{1}\} \qquad \forall t, t' \in T$

Additional restriction 2: If two consecutive time slots contain conflicting exams, then no conflicting exams scheduled in the next 3 time slots

Description

Add a new restriction which limits the number of consecutive time slots with conflicting exams to a maximum of 2, followed by 3 time slots without conflicting exams. The assumption, like for previous constraint, is that we are looking for conflicts only in consecutive time slots.
We use the same decision variable $\boldsymbol{k_{tt'}}$, and same constraints to link it to variable $\boldsymbol{y}$ as shown at page 10.

Constraints

- 2 consecutive time slots with conflicting exams must be always followed by 3 consecutive time slots without conflicting exams.
    - $\boldsymbol{k_{t+1,t+2} \leq 1 - k_{t,t+1}}$ $\quad \forall\, t \in \{1,2,\dots,|T|-2\}$
    - $\boldsymbol{k_{t+2,t+3} \leq 1 - k_{t,t+1}}$ $\quad \forall\, t \in \{1,2,\dots,|T|-3\}$
    - $\boldsymbol{k_{t+3,t+4} \leq 1 - k_{t,t+1}}$ $\quad \forall\, t \in \{1,2,\dots,|T|-4\}$
    - $\boldsymbol{k_{t+4,t+5} \leq 1 - k_{t,t+1}}$ $\quad \forall\, t \in \{1,2,\dots,|T|-5\}$

  With these constraints, in case we have 2 consecutive time slots having conflicting exams scheduled (e.g. t, t+1), then we force the next 3 consecutive time slots (t+2, t+3 and t+4) to not have conflicts with their adjacent time slots (conflicts checked between each time slots and his adjacent time slots).

Additional restriction 3: Bonus profit each time no conflicting exams are scheduled for 6 consecutive time slots

Description

Add a bonus profit each time there are 6 consecutive time slots without conflicting exams and use this bonus in the objective function.
We use the same decision variable $\boldsymbol{k_{tt'}}$ and same constraints to link it to variable $\boldsymbol{y}$ as shown at page 10.

Objective Function

$$\min \left(\sum_{\substack{e,e' \in E \\ e' > e \\ n_{ee'} > 0}} \sum_{\substack{t,t' \in T \\ |t-t'| \le 5}} \frac{2^{5-|t-t'|} \cdot n_{ee'} \cdot y_{ete't'}}{m} - \sum_{t \in \{1,2,\dots,|T|-5\}} b_t \right)$$

Decision Variables

- $\boldsymbol{b_t} = \mathbf{1}\ \forall\ t \in \{1, 2, \dots, |T| - 5\}$ if no conflicting exams are scheduled in time slots between $\boldsymbol{t}$ and $\boldsymbol{t} + \mathbf{5}$ included (every possible combination of time slots within the range is checked).

Constraints

- Each time no conflicting exams are scheduled for 6 consecutive time slots, a bonus score is added (also overlapping time periods are considered, e.g. no conflicting exams scheduled in time slots from t=1 to t=7 included, will result in 2 bonus points: 1 bonus point for time slots between t=1 and t=6, and another bonus point for time slots between t=2 and t=7).
  - $\boldsymbol{b_t} \ge \mathbf{1} - \sum_{\substack{t' \in \{t, t+1, \dots, t+4\} \\ t'' \in \{t'+1, \dots, t+5\}}} \boldsymbol{k_{t't''}} \qquad \forall\ t \in \{1,2, \dots, |T| - 5\}$
  - $\boldsymbol{b_t} \le \mathbf{1} - \frac{\sum_{\substack{t' \in \{t, t+1, \dots, t+4\} \\ t'' \in \{t'+1, \dots, t+5\}}} \boldsymbol{k_{t't''}}}{\mathbf{15}} \qquad \forall\ t \in \{1,2, \dots, |T| - 5\}$

  With these constraints, in case we have 6 consecutive time slots without conflicting exams, then $\boldsymbol{b_t}$ is forced to take value 1 thanks to first inequality.
  In case there's at least a conflict among the timeslots between t and t+5 ($\boldsymbol{k_{tt'}} = \mathbf{1}$), then $\boldsymbol{b_t}$ is forced to take value 0 thanks to second inequality (since $\boldsymbol{k_{tt'}}$ is a binary variable, 15 is an upper bound to the possible number of conflicts found by analyzing all possible combinations of timeslots between t and t+5, which are $\frac{6!}{4!2!}$).

- Domain constraints
  - $\boldsymbol{b_t} \in \{\mathbf{0}, \mathbf{1}\} \qquad \forall t \in T$

Additional restriction 4: At most 3 conflicting pairs of exams can be scheduled in same time slot

Description

Modify the original constraint that impose that no conflicting exams can be scheduled in the same time slot, to allow at most 3 conflicting pair of exams in the same time slot.

Constraints

- At most 3 conflicting pairs of exams can be scheduled in the same time slot.
    - $\sum_{\substack{e,e' \in E \\ e' > e \\ n_{ee'} > 0}} y_{ete't} \leq 3 \qquad \forall\, t \in T$

  With this constraint, we impose for each time slot the sum of all possible conflicting pairs for exams to be at most 3.

Conclusions

The aim of this chapter is to show the found results for the different formulations.

Base formulation results

In the case of base formulation, we were given benchmark values to be compared to our found results.

Name	Benchmark	Found results				
		Base formulation	Restriction 1	Restriction 2	Restriction 3	Restriction 4
Instance01	157,033	158,241	N.A.	N.A.	211,920	162,687
Instance02	34,709	46,627	N.A.	N.A.	N.A	56,141
Instance03	32,627	47,449	N.A.	N.A.	N.A	54,276
Instance04	7,717	11,422	N.A.	N.A.	N.A	15,700
Instance05	12,901	19,490	N.A.	N.A.	31,491	21,156
Instance06	3,045	N.A.	N.A.	N.A.	N.A	N.A.
Instance07	10,050	16,081	N.A.	N.A.	N.A	13,855
Instance08	24,869	30,629	N.A.	N.A.	38,255	28,572
Instance09	9,818	14,420	N.A.	N.A.	N.A	23,936
Instance10	3,707	N.A.	N.A.	N.A.	N.A	N.A.
Instance11	4,395	N.A.	N.A.	N.A.	N.A	N.A.

As we can see from the table above, the found results using the base formulation follow the given benchmark values, except for instances 6,10 and 11, for which the solver was not able to find a feasible solution within the 10 minutes set as parameter.
It has been tried to solve these instances changing the time limit to both 20 and 30 minutes but still no feasible solution was found.

When running the base formulation with an additional restriction between 1 and 2 (which involve new variables in the formulation), no feasible solutions were found within the time limits.

By adding restrictions 3 or 4, we obtain worse results most of the times (given the same time limit), even though we are not restricting the solution space. This could be due to the introduction of new variables in case of restriction 4 or due to the modification of a constraint for restriction 3 which now allows a bigger number of feasible solutions, thus increasing computational complexity and possibly requiring more time to reach the same results.

Equity measure variation results

We were not given any benchmark values to compare our results obtained from variations over the equity measures, therefore a toy test instance has been created to better analyze the behavior of the different equity measures.

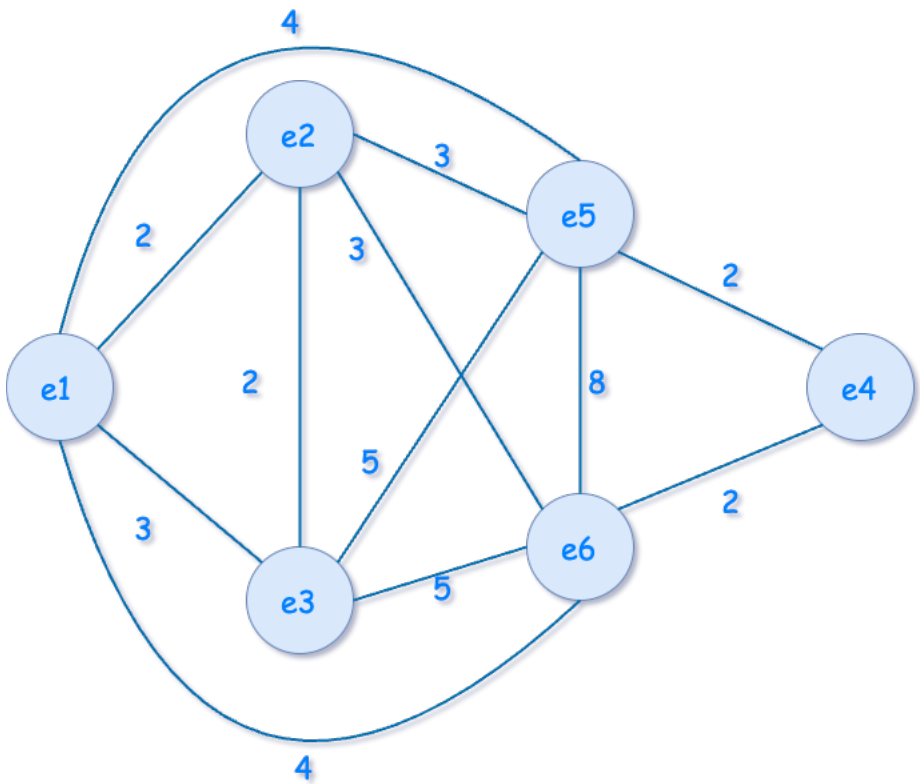
This toy instance has been created with the following input data:

Exams: {e1, e2, e3, e4, e5, e6}
Students: {s1, s2, s3, s4, s5, s6, s7, s8}

Enrollments table

Student	e1	e2	e3	e4	e5	e6
S1	✓	✓	✓		✓	✓
S2	✓		✓		✓	✓
S3				✓	✓	✓
S4			✓		✓	✓
S5	✓		✓		✓	✓
S6				✓	✓	✓
S7		✓	✓		✓	✓
S8	✓	✓			✓	✓

The conflict graph for the created toy instance is:



The different models behavior has been analyzed for different values of available time slots from 5 to 20 and one optimal solution for each combination is displayed in the following table:

Equity measure	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10	t11	t12	t13	t14	t15	t16	t17	t18	t19	t20
Available time slots: 5																				
Base	e5	e1	e3,e4	e2	e6															
1	e6	e5	e2	e3	e1,e4															
2	e5	e1	e6	e2	e3,e4															
3	e6	e5	e3	e2	e1,e4															
4	e6	e2	e1,e4	e3	e5															
Available time slots: 10																				
Base	e6			e1	e4	e3		e2		e5										
1	e5	e6		e2					e3	e1, e4										
2	e1			e3,e4		e2		e6		e5										
3	e6		e3			e1	e4	e2		e5										
4	e5		e6			e3		e4	e1	e2										
Available time slots: 15																				
Base	e6				e2		e4	e3			e1				e5					
1	e6	e5	e2											e3	e1,e4					
2	e1,e4		e5			e3							e2		e6					
3		e3,e4			e6			e2				e5			e1					
4	e5			e1				e3,e4			e2				e6					
Available time slots: 20																				
Base	e5						e6						e3				e2			e1,e4
1	e6	e5															e2		e3	e1,e4
2		e6			e2				e4		e3					e5			e1	
3	e1							e2	e4			e3				e5				e6
4	e6				e5						e2				e1,e4					e3

Even by analyzing a single optimal solution for each of the combinations of **equity measure-time slots available** is possible to notice how the different equity measures behave:

- **Base equity measure:** Whilst the other measures are only interested in whether the exams are conflicting or not, this measure considers also the number of conflicts that occur among exams (higher penalty if the number of enrolled students is higher). This can be seen by the fact that exams 5 and 6 (which have the highest number of conflicts) are always pushed to the extreme time slots of the examination period (or at least at a distance of 6 time slots, where the penalty term becomes ineffective), to maximize their distance and minimize the penalty term.
  This measure provides a fair schedule.
- **Equity measure 1:** this measure is based only on the average distance between any pair of conflicting exams. We can notice that this measure is not a good equity measure as the average distance is maximized when most of the exams are pushed towards the extreme time slots of the examination period, often scheduled in consecutive days.
  This is translated in some conflicting exams which are scheduled many days apart and other conflicting exams which are scheduled in back-to-back time slots, thus creating an imbalanced schedule.
- **Equity measure 2:** this measure is based on the minimization of the number of students with at least a pair of conflicting exams scheduled in consecutive time slots. As soon as there are enough time slots to guarantee that no conflicting exams are scheduled back-to-back, the measure does not make any distinction on how homogenously the exams are distributed in the examination period (see example with 15 time slots, where some conflicting exams had distance 2 and others had distance 6).
- **Equity measure 3:** this measure is based on the maximization of the minimum distance between any pair of conflicting exams, therefore it ensures that all conflicting exams are scheduled at least at that distance one from the other guaranteeing a quite fair schedule.
- **Equity measure 4:** this measure tries to solve the problems encountered with **equity measure 1** by taking into consideration also a MAD measure other than the average distance. The mean absolute deviation measure here is computed as the average of the absolute deviations of the distances of any two conflicting exams from total average distance. This measure seems to improve effectively **equity measure 1**, but the weights to be assigned in objective function are to be determined empirically.

Basically, every investigated equity measure is returning a quite fair schedule based on availability of time slots, each with a different focus.
The only exception is seen in **equity measure 1** which does not return a fair schedule.