Slack-based Robustness Estimators for the Curriculum-Based Course Timetabling Problem

Can Akkan · Ayla Gülcü · Zeki Kuş

Received: date / Accepted: date

Abstract Developing robust university course timetables is an important practical concern. Due to a variety of possible disruptions, i.e. changes in the input data affecting the constraints, quickly calculating an estimate of robustness to be used within a meta-heuristic, such as Simulated Annealing, would be very useful. In this research, we attempt to develop a set of slack-based estimators of a solution's robustness. To this end, we define 11 different slack measures (period, room and course-based) and use three summary statistics for each measure as an estimator of robustness. Preliminary experimental analysis of the performance of these estimators is done on a sample of 192 solutions for four International Timetabling Competition 2007 instances selected based on their diverse characteristics. The results suggest that a slack-based estimator can be used to identify a Pareto "band" rather than an approximate frontier that strikes a balance between probability of a solution on the true frontier being in the band and one not on the true frontier not being in the band.

Keywords Course timetabling · Robustness · Multi-criteria optimization

This work was supported by TÜBİTAK grant 217M475.

C. Akkan
Sabancı University, Sabancı Business School, Tuzla, 34956 İstanbul, Turkey.
Tel.: +90-216-4839685
E-mail: can.akkan@sabanciuniv.edu

A. Gülcü
Fatih Sultan Mehmet Univ., Dept. of Computer Science, Beyoğlu, 34445 İstanbul, Turkey.
Tel.: +90-212-5218100
E-mail: agulcu@fsm.edu.tr

Zeki Kuş
Fatih Sultan Mehmet Univ., Dept. of Computer Science, Beyoğlu, 34445 İstanbul, Turkey.
Tel.: +90-212-5218100
E-mail: zeki.kus@stu.fsm.edu.tr

1 Introduction

In a typical curriculum-based timetabling process at a university, an initial timetable, Σ_0, is prepared based on a set of constraints, which is then announced to the university staff, giving them some time to submit requirement changes, due to new constraints or data corrections. The timetable is then re-optimized taking these new constraints into account, while ensuring that the changes to Σ_0 is kept to a minimum. This second timetable is announced to the entire university and the students enroll in courses based on this timetable. Several types of disruptions that may affect this process are discussed in the literature (McCollum (2007), Müller et al (2005), Kingston (2013), Phillips et al (2017), Lindahl et al (2019)). As discussed in Phillips et al (2017) many different types of changes in data (disruptions) are possible before enrollment. Some new courses could be added, some others could be removed or canceled (see e.g. Yasari et al (2019)), some new faculty may arrive and some others may leave. Some disruptions could simply change feasibility of certain periods for some lectures. Some disruptions may affect either the availability or the capacity sufficiency of rooms for some lectures.

Problems in which constraints change over time are known as dynamic optimization problems, which fall into the category of optimization in uncertain environments. Meta-heuristics are quite often used for solving these problems (e.g. see Jin and Branke (2005) for a survey of evolutionary algorithms). Recently, there has been increasing interest in modeling and solving dynamic combinatorial optimization problems. One such problem closely related to timetabling is the graph coloring problem. Hardy et al (2018) develop heuristics for the dynamic graph coloring problem where edges are added/removed over time, randomly. They look into how information about the likelihood of future edge changes can be used to produce more robust colorings.

We say a timetable is *robust* if, when disrupted, its feasibility can be restored without significantly lowering its quality in terms of the objective function while keeping it relatively stable. We formulate the problem of identifying a robust timetable as a bi-criteria optimization problem where one objective is the quality of the solution measured as a function of the violated soft constraints (i.e., the penalty function), denoted by P, and the second one is a function that measures the robustness of the timetable, denoted by R.

We assume multiple number and types of disruptions can affect a given timetable. This makes calculating the robustness of a given solution quite time-consuming, since it requires optimally repairing that solution for a reasonably large sample of disruption scenarios. Thus, the main challenge of designing a meta-heuristic, such as Simulated Annealing (SA) to solve such a problem is designing an approximate measure of the robustness of a solution that can be calculated very quickly. In the work reported here, we develop and test some measures based on the degree and distribution of slack in a given timetable. The specific timetabling problem we address is the curriculum-based university course timetabling problem of ITC-2007 (see McCollum et al (2010)).

1.1 Disruption scenarios

Here we use the disruption scenarios that have been first developed and used by Akkan et al (2019). By assuming disruptions that affect two types of limited resources (time and rooms) we believe we introduce disruptions of sufficient variety and complexity. Specifically, we assume the following four types of disruptions may occur:

1. IP: The period to which a lecture of an instructor was assigned is no longer feasible for that instructor. This disruption is specified by a tuple $\langle i, p \rangle$ where i is an instructor, and p is the period to which one lecture of instructor i is assigned in Σ_0. If a disruption $\langle i, p \rangle$ is generated, unavailability constraints for all courses of instructor i at period p are added, unless such a constraint already exists.
2. CP: This disruption is specified by a tuple $\langle c, \mathcal{P}_1, \mathcal{P}_2 \rangle$ for course c. Given the set of periods available for course c, denoted by $\mathcal{P}_c^C$, $\mathcal{P}_1 \subseteq \mathcal{P}_c^C$ is a set of consecutive periods on the same day that become infeasible (unavailable) for course c and at least one of these periods is used by course c in Σ_0. $\mathcal{P}_2 \subseteq \mathcal{P} \setminus \mathcal{P}_c^C$ is a set of consecutive periods that become available for course c such that $|\mathcal{P}_2| \leq |\mathcal{P}_1|$.
3. CS: The number of students for a course is increased beyond the capacity of the room assigned to at least one lecture of that course. Note that this does not cause infeasibility, as room capacity is a soft constraint in ITC-2007, but increases the penalty of the initial timetable. This disruption is specified by a tuple $\langle c, s \rangle$, where s is the new number of students for course c. Even if some lectures of the course are currently assigned to rooms with enough capacity, all the lectures of this course are included in the set of room-disrupted lectures.
4. RP: Availability of one room is lost for one or two consecutive periods on the same day. This disruption is specified by $\langle r, p, d \rangle$, where p is the first period that room r becomes unavailable, and d is the number of periods that become unavailable.

A set of disruptions of these types is referred to as a *disruption scenario.* All disruptions in a given scenario are aggregated in two sets of disrupted lectures. The set of lectures e, whose assigned periods in Σ_0 become infeasible due to IP and CP disruptions is denoted by E^P (*period-disrupted* lectures). The set of lectures e, whose assigned rooms in Σ_0 become either infeasible due to RP disruptions or have insufficient capacity due to CS disruptions is denoted by E^R (*room-disrupted* lectures). Then, the set of disrupted lectures, E^D, equals $E^P \cup E^R$, with size δ.

1.2 The robustness measure

The robustness objective is expressed as minimizing $E(R(S, Y_S))$, the expected value of a disruption measure $R(S, Y_S)$, where S is a given solution and Y_S is the random variable representing the disruptions.

Let, $\mathcal{F}(\sigma_i)$ be the set of all solutions that are feasible with respect to a disruption scenario σ_i and $D(S_0, S_1)$ be the Hamming distance between assigned period arrays for all lectures $T_e(S_0)$ and $T_e(S_1)$ of these two solutions. $D(S_0, S_1)$ is equal to the sum, over all courses, of the number of lectures that are assigned to different periods in these two solutions. Then, we define the following neighborhood set for a given solution S_0 and disruption scenario σ_i with δ_i^p period and δ_i^r room-disrupted lectures:

$$\mathcal{N}(S_0, \sigma_i) = \{S : D(S_0, S) \leq f(\delta_i^p, \delta_i^r); S \in \mathcal{F}(\sigma_i)\} \tag{1}$$

Thus, if solution S_0 is disrupted by scenario σ_i, then switching to any solution in $\mathcal{N}(S_0, \sigma_i)$ would restore feasibility by rescheduling at most $f(\delta_i^p, \delta_i^r)$ lectures to a different period, where $f : (\mathbb{N}, \mathbb{N}) \to \mathbb{N}$. Then, we define the robustness measure for a given solution S_0 and a disruption scenario σ_i as,

$$R(S_0, \sigma_i) = \min_{S \in \mathcal{N}(S_0, \sigma_i)} \Big(P_{ave} \cdot 1_{D(S,S_0) > \delta_i^p} + (P(S) - P(S_0))^+ \Big) \tag{2}$$

where $x^+ := \max(0, x)$ and P_{ave} is the average per lecture penalty for a randomly generated sample of solutions (see Gülcü and Akkan (2020)). P_{ave} is an additional penalty term added so that solutions that only reschedule period-disrupted lectures to different periods are favored. Thus, in addition to quality robustness measured by $(P(S) - P(S_0))^+$, by adding a fixed penalty cost for rescheduling more lectures than the period-disrupted lectures, $R(S_0, \sigma_i)$ incorporates a measure of solution stability. Solution stability is further ensured by adding the constraint $D(S_0, S) \leq f(\delta_i^p, \delta_i^r)$ in defining $\mathcal{N}(S_0, \sigma_i)$. If $\mathcal{N}(S_0, \sigma_i) = \emptyset$, then $R(S_0, \sigma_i)$ is set to a large value, denoted by B.

For solution S, an estimate of $E(R(S, Y_S))$ is calculated as a sample average $\frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}} R(S, y)$ for a sampled set of disruption scenarios $\mathcal{Y}$, since a closed-form calculation of $E(R(S, Y_S))$ is not possible. Given the robustness measure, R, and a set of randomly generated sample of disruption scenarios, $\underline{\sigma} = \{\sigma_1, \sigma_2, \ldots, \sigma_N\}$, $E(R)$ is estimated by $\overline{R}(S, \underline{\sigma}) = (1/N) \sum_{i=1}^{N} R(S, \sigma_i)$. $\overline{R}(S, \underline{\sigma})$ is taken as the true robustness measure, using a reasonably large N. This approach bears some resemblance to the Sample Average Approximation (SAA) method of Kleywegt et al (2002). The algorithm that is used to repair solution S subject to a given disruption scenario is the Simulated Annealing algorithm discussed in detail in Akkan et al (2019), in which $f(\delta_i^p, \delta_i^r) = 2 \times \delta_i^p + 0.25 \times \delta_i^r$, and $B = 1200$.

2 Slack-based Estimators

The tested estimators are summary statistics of some measures of slack in a given timetable. These measures can be classified into three groups. The first group provides measures of slack for each period. Let, $X(p, r)$ equals 1 if room r is used at period p by some lecture, 0 otherwise; and $Rm(p)$ equals the number of rooms used at period p, while the total number of rooms is Rm.

Then, assuming the rooms are indexed in increasing capacity, define $\mathbf{RSA}[p]$, room-based slack at period p, as

$$\mathbf{RSA}[p] = \begin{cases} \frac{1}{Rm(p)} \sum_r \sum_{q>r} 1_{([X(p,r)=1, X(p,q)=0)}, & \text{if } Rm(p) > 0 \\ (Rm-1), & \text{if } Rm(p) = 0 \end{cases} \tag{3}$$

The summation term in Eqn. 3, $\rho(p,r) = \sum_{q>r} 1_{([X(p,r)=1, X(p,q)=0)}$ gives for every room used at a period, the number of available larger capacity rooms in the same period.

It is quite reasonable to have decreasing marginal benefit in increasing $\rho(p,r)$, so an alternative slack measure could make use of an exponential utility function as $\mathbf{RSU}[p] = \frac{1}{Rm(p)} \sum_r \sum_{j>=1}^{\rho(p,r)} e^{-j}$ for a given period p.

Alternatively, we can assume there is utility in having at least one larger capacity room for a lecture scheduled at a given period, say at (p,r). In that case, we would have $\rho(p,r) > 0$. Then, another slack measure can be defined as $\mathbf{RSB}[p] = \frac{1}{Rm(p)} \sum_r 1_{\rho(p,r)>0}$ for a given period p.

The second group of estimators make use of slack measured for each room. For room r, letting, $\pi(r) = \sum_t 1_{X[t,r]=0}$ denote the the number of available periods in the same room, we define $\mathbf{PSU}[r] = \sum_{j=1}^{\pi(r)} e^{-j}$ as a slack measure for a given room r.

The next measure is a more finely grained version of $\mathbf{PSU}[r]$, where the periods are sub-divided into daily sets. Let the day of a given period p be denoted by $D(p)$. Then, $\xi(d,r) = \sum_{p:D(p)=d} 1_{X[p,r]=0}$ represents the number of available time-slots on day d at room r, and we define $\mathbf{DSU}[d,r] = \sum_{j=1}^{\xi(d,r)} e^{-j}$ as a slack measure for a given room r on day d.

The third group of estimators measure course-specific availability of periods. We first let,

$\mathcal{Y}(p)$	= the set of courses scheduled at period p.
$\mathcal{FR}(p)$	= the set of free rooms in period p
$FR(p)$	= the number of free rooms at period p.
$K(r)$	= the capacity of room r
$K^{max}(p)$	$= \max_{r \in \mathcal{FR}(p)} \{K(r)\}$
$S(c)$	= the number of students planned for course c
$\mathcal{F}(c)$	= the set of feasible periods for course c.
$\mathcal{C}(c)$	= the set of conflicting courses for course c

Note that courses in the same curriculum or taught by the same teacher are referred to as conflicting courses. We then define the following sets,

$$\mathcal{AP}(c) = \{p : p \in \mathcal{F}(c), FR(p) > 0\}$$
$$\mathcal{AP}^+(c) = \{p : p \in \mathcal{AP}(c), K^{max}(p) > S(c)\}$$
$$\mathcal{CP}(c) = \{p : p \in \mathcal{AP}(c), \mathcal{Y}(p) \cap \{\mathcal{C}(c) \cup c\} = \emptyset\}$$
$$\mathcal{CP}^+(c) = \{p : p \in \mathcal{AP}^+(c), \mathcal{Y}(p) \cap \{\mathcal{C}(c) \cup c\} = \emptyset\}$$

Thus, $\mathcal{AP}(c)$ gives the set of periods available for course c, $\mathcal{CP}(c)$ gives the set of conflict free periods available for course c and $\mathcal{CP}^+(c)$ gives the set of conflict free periods available for course c, having at least one free room with sufficient capacity.

Given these sets, we can define the array $\mathbf{C}[c]$ that contains the number of conflict-free available periods for each course, as $\mathbf{C}[c] = |\mathcal{CP}(c))|$. Similarly, the array of the number of conflict-free available periods with free rooms of sufficient capacity, for each course c, is defined as $\mathbf{R}[c] = |\mathcal{CP}^+(c)|$.

Rather than simply counting the number of available periods, as a basis of the next slack measure, we calculate the utility of the conflict-free available periods using an exponential utility function, that gives decreasing marginal utility with increasing number of conflict-free available periods. The array of these utilities is defined as $\mathbf{UC}[c] = \sum_{j=1}^{|\mathcal{CP}(c)|} e^{-j}$, for each course c. Similarly, we define a utility function that is based on conflict-free available periods with available rooms of sufficient capacity as $\mathbf{UR}[c] = \sum_{j=1}^{|\mathcal{CP}^+(c)|} e^{-j}$, for each course c. Furthermore, one can argue that the value of $\mathbf{C}[c]$ depends on the number of lectures of course c, $L(c)$, so we defined two additional arrays $\mathbf{CED}[c]$ and $\mathbf{CER}[c]$, as $\mathbf{CED}[c] = \mathbf{C}[c] - L(c)$, and $\mathbf{CER}[c] = \mathbf{C}[c]/L(c)$.

Given these eleven slack measuring arrays (3 defined for each period, 2 for each room, and 6 for each course), we calculate the average, standard deviation and the coefficient of variation (standard deviation over average) of each array as estimators of the robustness of the given timetable. These three summary statistics for a given slack measure $\mathbf{S}$ are denoted as $\overline{S}$, SD_S, CV_S, respectively.

3 Computational results

For the computational experiments we selected four ITC-2007 instances, namely ITC1, ITC2, ITC5 and ITC12. They are among the most constrained (thus potentially difficult) instances in terms of conflict intensity, teacher availability, and room occupancy (Bonutti et al (2012)). Then, for each instance, a set of 48 solutions were selected to carry out correlation analysis between the slack metrics and the robustness measure, $\overline{R}$. The purpose of the selection procedure was to obtain a diverse set of solutions, from among the set of solutions accepted through a Simulated Annealing algorithm designed to minimize the penalty. A brief discussion of the selection procedure is provided below, interested readers could find the details in Akkan et al (2020).

3.1 Solutions selected for analysis

The solutions that were found in the SA search process were used to form a network of solutions, where each solution is represented by a node. Selection of solutions for the computational experiments was based on two of their characteristics: the penalty value and the degree of the node. For each ITC instance four networks were generated, and 12 solutions were selected from each network. For the instance ITCi, the network $N_i^{nc,s}$ was generated by collecting *nc thousand* solutions accepted by the SA algorithm (we used 50 and 100 thousand). Starting with the solution accepted in the last iteration of SA, going backwards and skipping every s^{th} accepted solution, a total of

Table 1 Penalty and degree intervals used to select the sample of solutions from $N_2^{50,0}$

P : 75	76	77	78	79	80
[13, 32]	[13, 32]	[13, 32]	[33, 52]	[33, 52]	[53, 72]
[173, 192]	[133, 152]	[113, 132]	[93, 112]	[93, 112]	[73, 92]

nc thousand solutions are collected into the network (we used s equals 0 and 1). Letting $G(N, E)$ denote a solution network, we let the neighbors of node v be the set of nodes w such that $D(v, w) \leq \rho$, where $D(v, w)$ is the number of lectures with different assigned periods in v and w, and ρ is an integer parameter. ρ could be seen as the maximum number of events that would need to be rescheduled to a *different period* in order to respond effectively to a disruption scenario. We calculated ρ for each ITC instance as $2 \times N^p + 0.25 \times N^r$, where N^p and N^r are the maximum number of period-based and room-based disruptions possible for that instance (for ITC2 $N^p = 8$ and $N^r = 16$ events). Thus, the degree of a solution is likely to be correlated with the robustness of the solution, as it indicates the size of the solution space that could be used to repair that solution.

Frequency tables were formed of all solutions in each network based on intervals of penalty and degree, and then a solution was selected from the solutions that fall into selected intervals. The selections were made from among the solutions with penalties that are close to the minimum penalty value, P_{min}, found by the SA algorithm. For each network, six penalty-intervals were selected. For example, for ITC2 $P_{min} = 75$ and all solutions had one of the six penalty values listed in Table 1, so intervals were only defined for the degrees. Given the penalty interval (or the penalty) we randomly sampled one solution from the smallest degree interval, and the second solution from the largest degree interval. In a few cases when a degree interval contained only already selected solutions, we moved to the adjacent degree interval for the same penalty interval.

3.2 Correlation analysis

Pearson correlations coefficients, ρ_i^m, were calculated between each slack metric, say m, and the robustness measure $\overline{R}$ using the set of 48 solutions for each instance $ITCi$. Then the absolute values of these correlations were ranked among those for each instance, in decreasing order so that the largest one is ranked first. Letting ϱ_i^m denote the rank of $|\rho_i^m|$, the average rank of metric m was calculated as, $\overline{\varrho}^m = 1/4 \sum_{i \in \{1,2,5,12\}} \varrho_i^m$. For the nine best ranking metrics, their correlation coefficients and average ranks, $\overline{\varrho}^m$, are reported in Table 2. For each of these correlation coefficients, a test of hypothesis was done where the null hypothesis states the correlation is equal to zero.

Based on the correlations presented in Table 2 we chose three of the metrics for further analysis, namely, CV_{CER}, $\overline{UC}$, and CV_C, which are the best ones for their corresponding arrays. CV_{CER} has the best overall average rank with

Table 2 Pearson correlation coefficients with $\overline{R}$

	CV_{CER}	$\overline{UC}$	$\overline{CER}$	CV_{UC}	SD_{UC}	CV_C	$\overline{RSU}$	$\overline{DSU}$	CV_{RSA}
Ave. Rank	8.5	9.5	10.25	10.75	11.75	12.25	13.25	14	14.25
ITC1	-0.226	-0.227	-0.184	0.211	0.205	0.142	0.106	$0.285^{\flat}$	-0.120
ITC2	$0.292^{\flat}$	$-0.384^{\sharp}$	-0.060	$0.409^{\sharp}$	$0.411^{\sharp}$	$0.303^{\flat}$	$0.429^{\sharp}$	0.088	$-0.426^{\sharp}$
ITC5	$0.296^{\flat}$	-0.169	$-0.321^{\flat}$	0.160	0.150	0.256^{*}	-0.219	0.199	0.152
ITC12	$0.290^{\flat}$	$-0.319^{\flat}$	$-0.375^{\sharp}$	0.271^{*}	0.255^{*}	0.277^{*}	0.176	0.154	-0.205

$\sharp$: $p < .01$, $\flat$: $p < .05$, $*$: $p < .10$

statistically significant correlations for three instances. Correlation between $\overline{UC}$ and $\overline{R}$ is negative for all four instances, although for two of them we have sufficiently low p-values to reject the null hypothesis of 0 correlation. CV_C, on the other hand, is consistently positively correlated with $\overline{R}$ and for three of the instances, the correlations associated with CV_C are different from zero at a statistically significant level. Recall that $\mathbf{C}[c]$ array contains the number of conflict-free available periods for each course. Thus, the positive correlation associated with CV_C suggests that the more evenly such periods are distributed across courses and the larger the average number of conflict-free available periods is, the more robust the solution would be (with smaller $\overline{R}$). CV_{UC}, the utility-adjusted version of CV_C, yields a similar performance to CV_C. The negative correlations associated with $\overline{UC}$ is consistent with this interpretation.

For all instances, the scatter plots of the nine slack metrics with $\overline{R}$ were plotted. We observed for ITC5 that in the plot for $\overline{RSU}$ there is an outlier solution that is increasing the correlation, however the hypothesis testing resulted in not rejecting a zero correlation (see Figure 1). For the other scatter plots, we did not see such a case of a single outlier.

3.3 Accuracy in identifying the Pareto frontier

The planned use of a slack metric is as an estimate of robustness within a multi-objective Simulated Annealing algorithm (MOSA) with two objectives: the penalty of the solution and its robustness. A MOSA algorithm maintains an archive of solutions which comprises the best Pareto frontier at each iteration. A new solution enters the frontier if there is no solution in the current frontier which dominates it. The number of solutions that dominate a given solution s is referred to as the domination count of solution s and the rank of that solution, $r(s)$, is equal to its domination count plus 1. So, a frontier is comprised of solutions of rank 1. Since the MOSA algorithm will be designed to use the slack metric M rather than the robustness measure $\overline{R}$, and M is a estimator, a solution with rank 2 defined by (P, M) might easily be on the Pareto frontier defined by $(P, \overline{R})$. Thus, it would be reasonable to keep in the archive, solutions s with $r(s) \leq K$, where K is an integer cutoff value, rather than $r(s) = 1$.

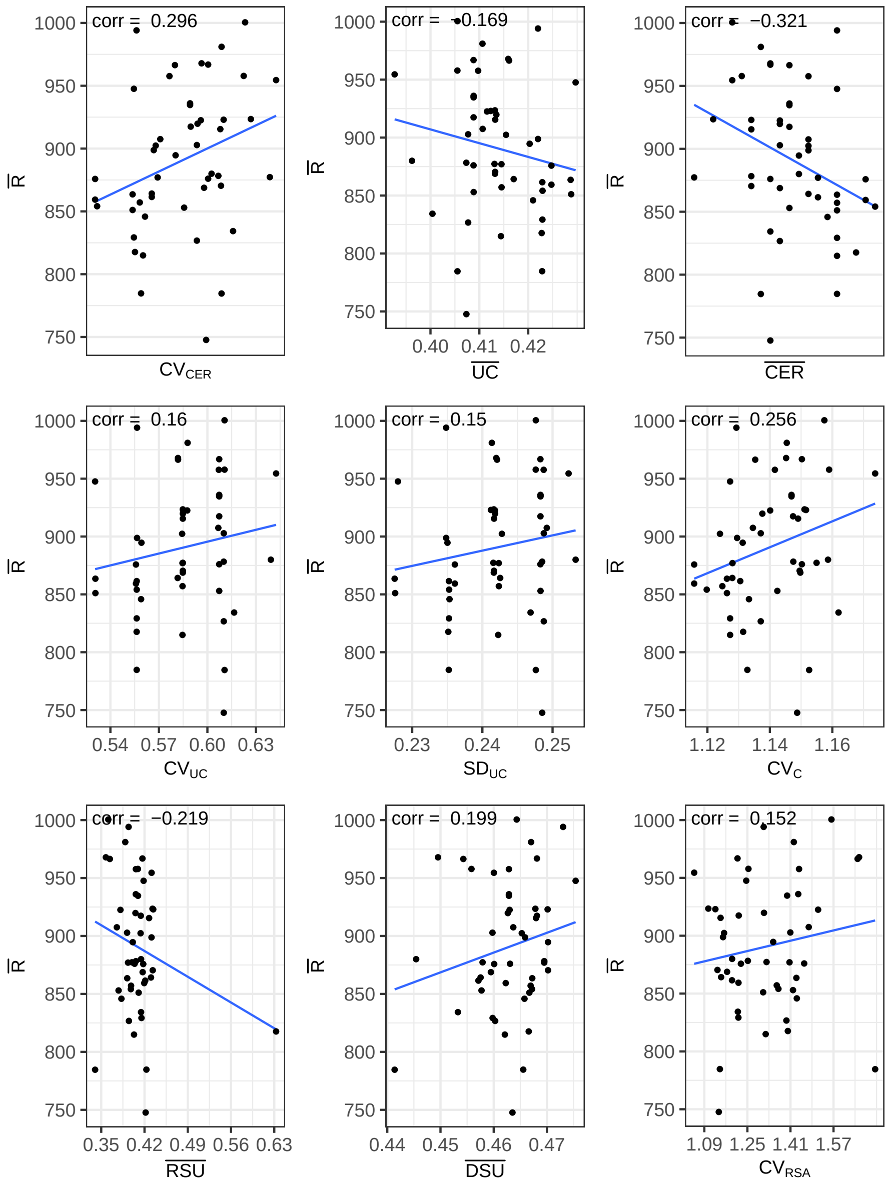

Fig. 1 ITC5: scatter plots of $\overline{R}$ versus selected slack metrics

In this case, the final archive at the end of the MOSA algorithm's run, $\mathcal{A}$, would be a short-list of solutions that are highly likely to contain the solutions forming the Pareto frontier based on $(P, \overline{R})$, $\mathcal{F}$. We then would calculate the robustness $\overline{R}(s)$ for all $s \in \mathcal{A}$ and obtain the approximate Pareto frontier $\widetilde{\mathcal{F}}$, based on $(P, \overline{R})$.

Given the approach discussed above, for a metric M to performs well, we would want a large $P(s \in \mathcal{A}|s \in \mathcal{F})$, and similarly a large $P(s \notin \mathcal{A}|s \notin \mathcal{F})$. To estimate these probabilities, we have done the following experimental analysis. For all 48 solutions of each ITC instance i we calculated their $\overline{R}$ values and determined the corresponding Pareto frontier $\mathcal{F}_i$. Then for each slack metric M, we calculated the rank $r(s)$ of each solution based on (P, M) and determined the solutions that fall into the archive $\mathcal{A}(K)_i$ defined by the cutoff value K. Based on the combined sample of 192 solutions, $\mathcal{S}_i$ for $i = 1, 2, 5, 12$, the estimates of $P(s \in \mathcal{A}|s \in \mathcal{F})$ and $P(s \notin \mathcal{A}|s \notin \mathcal{F})$, denoted by $f^+(K)$ and $f^-(K)$, respectively, are defined as:

$$f^+(K) = \frac{\sum_i |\{s : s \in \mathcal{A}(K)_i \ , s \in \mathcal{F}_i\}|}{\sum_i |\mathcal{F}_i|} \tag{4}$$

$$f^-(K) = \frac{\sum_i |\{s : s \in \mathcal{S}_i \setminus \mathcal{A}(K)_i \ , s \in \mathcal{S}_i \setminus \mathcal{F}_i\}|}{\sum_i |\mathcal{S}_i \setminus \mathcal{F}_i|} \tag{5}$$

Table 3 Accuracy in identifying the solutions on the Pareto frontier

	$\overline{UC}$		CV_{CER}		CV_C	
K	$f^+(K)$	$f^-(K)$	$f^+(K)$	$f^-(K)$	$f^+(K)$	$f^-(K)$
1	0	0.939	0	0.939	0.091	0.917
2	0.182	0.884	0.273	0.884	0.273	0.884
3	0.455	0.840	0.364	0.823	0.364	0.818
4	0.545	0.796	0.455	0.762	0.636	0.768
5	0.636	0.746	0.636	0.729	0.727	0.740
6	0.636	0.713	0.636	0.685	0.727	0.685
7	0.636	0.669	0.727	0.641	0.818	0.646
8	0.727	0.608	0.727	0.613	0.818	0.597

Table 3 presents the above fractions for the three slack metrics selected based on the correlation analysis in Section 3 for the cutoff values $K = 1, \ldots, 8$. For $K \geq 4$, we can see that $f^+(K)$ is consistently larger for CV_C than for the two other metrics. On the other hand, $f^-(K)$ is consistently larger for $\overline{UC}$ than the other two. The chart in Figure 2 suggests that $K = 5$ constitutes a good trade-off between $f^+(K)$ and $f^-(K)$ for CV_C. At $K = 5$, CV_C gives a better performance than CV_{CER} in terms of both $f^+(K)$ and $f^-(K)$. On the other hand, comparing CV_C and $\overline{UC}$ we observe that their $f^-(K)$ values are almost identical but CV_C has a better $f^+(K)$.

4 Concluding remarks

In this work an attempt has been made to develop approximate measures of robustness in the form of slack-based estimators that could be used within a Simulated Annealing algorithm, which would identify the Pareto frontier

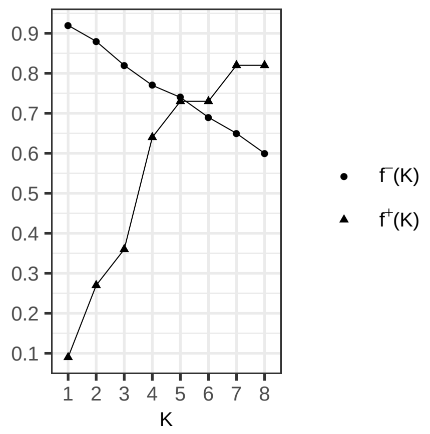

Fig. 2 CV_C accuracy: the tradeoff between $f^+(K)$ and $f^-(K)$

defined by the penalty of a timetable and a measure of its robustness. Finding such fast-to-compute estimators is needed because the calculation of the robustness measure requires too much computational effort when there are multiple disruptions of different types.

The approximate nature of the estimators suggest that they can be used to identify a Pareto "band", comprised of solutions with rank less than or equal to a cutoff value, as opposed to the frontier comprised of solutions with rank equals 1. The performance quality of a band is judged by how accurately it distinguishes solutions that are on the true Pareto frontier (defined by the robustness measure) from those that are not. To this end, for each cutoff value and estimator, we calculated estimates of the probability of a solution on the true frontier being in the band and one not on the true frontier not being in the band. This is done on a sample of 192 solutions (48 solutions for each of four ITC 2007 instances). The results suggest that CV_C, the coefficient of variation of the number of conflict-free available periods for each course, is the best one among the 33 estimators tested.

The experimental analysis presented here should be seen as a preliminary work, as this is currently a work in progress. We are implementing a MOSA (Multi-Objective Simulated Annealing) algorithm that uses a given estimator and maintains the Pareto "band" as opposed to the Pareto frontier in its archive of solutions. Potential extensions of this work could include finding other slack-based estimators, and also investigating different robustness measures for which slack-based estimators can provide a better performance.

References

Akkan C, Gülcü A, Kuş Z (2019) Minimum penalty perturbation heuristics for curriculum-based timetables subject to multiple disruptions, unpublished

Akkan C, Gülcü A, Kuş Z (2020) Search space sampling by simulated annealing for identifying robust solutions in course timetabling. In: Proceedings of the 2020 IEEE Congress on Evolutionary Computation, to appear

Bonutti A, De Cesco F, Di Gaspero L, Schaerf A (2012) Benchmarking curriculum-based course timetabling: formulations, data formats, instances, validation, visualization, and results. Annals of Operations Research (194):59–70, DOI 10.1007/s10479-010-0707-0

Gülcü A, Akkan C (2020) Robust University Course Timetabling Problem Subject to Single and Multiple Disruptions. European Journal of Operational Research 282:630–646, DOI 10.1016/j.ejor.2019.11.024

Hardy B, Lewis R, Thompson J (2018) Tackling the edge dynamic graph colouring problem with and without future adjacency information. Journal of Heuristics 24(3):321–343, DOI 10.1007/s10732-017-9327-z

Jin Y, Branke J (2005) Evolutionary optimization in uncertain environments-a survey. IEEE Transactions on Evolutionary Computation 9(3):303–317, DOI 10.1109/TEVC.2005.846356

Kingston JH (2013) Educational Timetabling. In: Uyar AS, Ozcan E, Urquhart N (eds) Automated Scheduling and Planning: From Theory to Practice, Springer, Berlin, Heidelberg, DOI 10.1007/978-3-642-39304-4

Kleywegt AJ, Shapiro A, Homem-de Mello T (2002) The Sample Average Approximation Method for Stochastic Discrete Optimization. SIAM Journal on Optimization 12(2):479–502, DOI 10.1137/S1052623499363220

Lindahl M, Stidsen T, Sørensen M (2019) Quality recovering of university timetables. European Journal of Operational Research 276(2):422–435, DOI 10.1016/j.ejor.2019.01.026

McCollum B (2007) A perspective on bridging the gap between theory and practice in university timetabling. Practice and Theory of Automated Timetabling, LNCS 3867:3–23

McCollum B, Schaerf A, Paechter B, McMullan P, Lewis R, Parkes AJ, Di Gaspero L, Qu R, Burke EK (2010) Setting the research agenda in automated timetabling: The second international timetabling competition. INFORMS Journal on Computing 22(1):120–130, DOI 10.1287/ijoc.1090.0320

Müller T, Rudová H, Barták R (2005) Minimal Perturbation Problem in Course Timetabling. In: Burke E, Trick M (eds) Practice and Theory of Automated Timetabling V. Lecture notes in computer science (Vol. 3616), vol 3616, pp 126–146, DOI 10.1007/11593577_8

Phillips AE, Walker CG, Ehrgott M, Ryan DM (2017) Integer programming for minimal perturbation problems in university course timetabling. Annals of Operations Research 252(2):283–304, DOI 10.1007/s10479-015-2094-z

Yasari P, Ranjbar M, Jamili N, Shaelaie MH (2019) A two-stage stochastic programming approach for a multi-objective course timetabling problem with courses cancelation risk. Computers & Industrial Engineering 130:650–660, DOI 10.1016/j.cie.2019.02.050