Bare Demo of IEEEtran.cls for IEEE Journals

Angus Kenny[a], Dhananjay Thiruvady[b], Davaatseren Baatar[c], Andreas T. Ernst[c], Xiaodong Li[d], Mohan Krishnamoorthy[e], Gaurav Singh[f]

***Abstract*—Exact methods are useful when solving combinatorial optimisation problems, although their time complexity is often high. Conversely, meta-heuristics are adept at finding feasible solutions in reasonable time, but their stochastic nature can make it hard to refine solutions effectively. To this end, a class of algorithms known as matheuristics was developed, which harness the strengths of both these methods.**

This paper presents a matheuristic called merge search, an iterative decomposition algorithm which creates a reduced sub-problem by grouping variables that share common values across a population of solutions. Merge search can be thought of as a generalisation of crossover, from evolutionary algorithms; instead of combining two solutions to create a single offspring solution, the sub-space that is spanned by the entire population is searched for the optimal way to combine all solutions.

Merge search is applied to two very different NP-Hard problems: the constrained pit limit (CPIT) problem, from the field of open-pit mining; and the well-studied Steiner tree problem in graphs (STPG). The CPIT results are used to show the effectiveness of merge search, improving the best-known bounds in two of the six instances tested; while the STPG results demonstrate the versatility of this method.

***Index Terms*—IEEE, IEEEtran, journal, LaTeX, paper, template.**

I. Introduction

COMBINATORIAL optimisation problems, at their very basis, involve searching a finite (but often *very* large) and discrete search space to find an optimal object [1]. Optimisation problems modelling situations in the real-world can be very messy and complex, often involving many variables and constraints [2]. Traditional mathematical solvers are not well-suited to these messier, large-scale problems; making finding ways to decompose complex problems into smaller, more manageable sub-problems an important topic of study [3].

There are several factors that can contribute to how difficult a problem is to solve [4]. The first being the number of decision variables that need to be considered. The second factor is the interaction between decision variables. If a problem is completely separable, it can be solved by completely decomposing it into its individual variables and optimising them separately. The other extreme to this is when it is completely inseparable, meaning the entire problem must be solved at once. Most problems tend to fall somewhere between these two ends of the spectrum, making a search for an effective way to ascertain these interactions an important area of research [5], [6]. Finally (although there are many others), the number, and type, of constraints are also a factor that can contribute to a problem's complexity. Constraints can take many different forms, from production limits [7]; to precedence constraints, where a variable may not be selected until all variables from a specified set are also selected [8]; to constraints on the types of values the variables can take — restricting decision variables to integer values can make a problem much harder than if they were real-valued [9].

The most effective method of solving large scale problems is by a method called *divide-and-conquer* [**?**]. In this technique, a large problem is decomposed into several smaller problems with fewer variables or constraints, which can then be solved and the information from the solutions to these sub-problems used to inform the over-all global search. There are many ways to decompose problems to make them more manageable. One way is to divide the variables into subsets and solve a version of the main problem on each subset of variables and then perform some repair operation to "stitch together" the results of all the sub-problems [**?**]. Another way is to start with a small subset of variables and solve the problem for those variables, and iteratively add more and more variables to the main problem by solving smaller sub-problems to choose them [**?**]. Still another way is to aggregate the variables into a number of groups that each act as a single variable in the sub-problem, which is then solved and the groups iteratively disaggregated to achieve better quality solutions [**?**]. Whichever method is chosen, it is not trivial to decide the best way to decompose any problem; so an efficient method of automatically determining an effective problem decomposition would be very valuable to the research community at-large.

Another major issue that arises when solving large scale combinatorial problems is the problem of global vs. local search [**?**]. The exponential size of the search space and the intricate interactions between decision variables mean that the fitness landscape for many problems can be extremely complex, with many local optima and basins of attraction that must be escaped in order to find the global optimum. This complex fitness landscape and integer values for variables mean that most gradient-based search methods are not very effective tools for solving these problems.

Population-based meta-heuristic algorithms are particularly adept at performing global search, as they use various techniques to intelligently sample a wide region of the search

The authors are with: [a]School of Engineering and Information Technology, University of New South Wales, Canberra ACT, Australia; [b]School of Information Technology, Deakin University, Geelong VIC, Australia; [c]School of Mathematics, Monash University, Clayton VIC, Australia; [d]School of Science, RMIT University, Melbourne VIC, Australia; [e]School of Information Technology and Electrical Engineering, The University of Queensland, St Lucia QLD, Australia; [f]BHP, Perth WA, Australia.

space. This wide-ranging sampling allows them to identify areas of the search space that are the most promising and helps them to avoid getting caught in local optima; however, having identified these promising regions, they are typically not very good at producing refined solutions from within them [**?**]. Conversely, mathematical programming methods are very poor at global search, as the large number of decision variables and constraints make the problems intractable; but once a good area of the search space has been identified, and the problem can be reduced to the variables and constraints only concerning this area, these techniques are very good at finding the locally optimum solution.

It is this trade-off between exploration and exploitation that is the basis for the family of algorithms called *hybrid meta-heuristics* [**?**], [**?**]. These methods typically combine a meta-heuristic, which performs general global search or decomposes the problem into a reduced sub-problem; and a mathematical programming solver which provides refined solutions to the sub-problems that the meta-heuristic produces. This interplay between the two elements goes back-and-forth, iteratively improving the solutions they are producing by harnessing the strengths of both kinds of algorithms. These techniques have become a subject of a great deal of research lately [3], as they have been shown to be effective on many different types of problems; and a good way to address a lot of the issues encountered when attempting to solve complex, large scale problems.

Finally, many techniques in the literature are very good at solving a particular problem, but cannot be generalised to solve other problems. These techniques often rely on domain knowledge of the problem having being "hard-coded" into the problem model, or on some customised representation of the problem that does not translate well to another domain. As the solution to most combinatorial optimisation problems can be represented as a binary string, there is opportunity for a technique that operates purely on these basic representations to generalised into a framework for solving any problem of this type, irrespective of the domain.

IN the mining industry in Australia, open-pit mining is a problem of significant interest [10]. The planning and production scheduling associated with mines can lead to significant cost savings and profits for the operators of the mines. In particular, determining the value of a mine and the areas of excavation in the shortest possible time-frame can lead to very large gains [**?**].

Extracting and processing the ore from the pits is the main focus of open pit mining. Specifically, the order in which materials are extracted and processed can lead to significant profits and savings [**?**]. In determining the order, there are several constraints that must be satisfied, including precedences between blocks and resource limits. Hence, the open-pit block scheduling problem is referred to as the *precedence constrained production scheduling problem* (PCPSP) [**?**], [**?**].

The PCPSP requires identifying a schedule to extract blocks in a pit that maximises the net present value (NPV) of the blocks over time. Each block is associated with a positive value (profit) or a negative value (cost) and if mined, it is either processed or discarded (specified as destinations). The mining of blocks is subject to resource and precedence constraints. Resource limits apply to the amount of ore mined and processed during a period, while precedences between blocks exist due to pit slope constraints. That is, in order to reach certain blocks, other blocks on top of them need to be extracted and processed or discarded first.

The PCPSP can be formulated as a Mixed Integer Program (MIP). For real world instances, solving the MIP or even the linear programming (LP) relaxation of the problem is challenging due to the large number of variables and constraints. For example, the problem instances currently available in the literature can be very large with nearly 100,000 blocks and over a million precedence constraints. These need to be replicated over multiple time periods resulting in MIP formulations with an excess of 10 million constraints.

Finding ways to solve large MIPs with reasonable computational resources has been given attention recently. The methods include decomposition approaches, hierarchical methods and MIP-based large neighbourhood search. Among MIP-based decompositions, Lagrangian relaxation [**?**], column generation [**?**] and Benders' decomposition [**?**] have been widely applied and proved very effective. Other approaches for efficiently solving MIPs is to identify aggregations of variables so that a problem of reduced size can be solved [**?**], [**?**], [**?**]. Another class of methods are based on a large neighbourhood search (LNS) [**?**]. The main idea underlying these methods is to start with a feasible solution, identify a neighbourhood (possibly a very large), and find efficient ways to search the neighbourhood. When LNS is combined with MIPs, the resulting methods have proved very effective [**?**], [**?**].

The main contribution of this study is a novel matheuristic algorithm, called *Merge Search*. It combines concepts from LNS, genetic algorithms [**?**] and the efficiency of solving MIPs. However, it is substantially different to any of these or other existing methods in two main aspects. First, the neighbourhood leading to the restricted MIP is obtained from an aggregation of variables in original model using a population of solutions (potentially very large populations). Using different input parameters (e.g. population size), the solving time of the restricted MIP can be systematically controlled. Second, Merge Search is particularly designed for parallel or distributed computing, thus allowing additional computing resources to be used effectively in tackling these challenging problems. Hence, a second contribution of this study is develop and investigate parallel implementations of Merge Search via the Message Passing Interface (MPI) [**?**]. A third contribution of this study is a Branch & Bound method built around the LP relaxation of the problem. This is intended to show that even though the LP relaxation of the large MIPs can be solved relatively efficiently, the associated problems cannot be solved to optimality by a standard branch and bound approach. Despite this, we find that this method leads to identifying the best known upper bounds for benchmark instances of open pit mining.

Recently, a method which uses a population of solutions to identify a search space for MIP model is construct, solve, merge and adapt (CMSA) [**?**], [**?**], [**?**], [**?**], [**?**]. [**?**] show that CMSA can be effectively applied to the minimum common

string partition and minimum covering arborescence problems. [**?**] investigate CMSA for the repetition-free longest common subsequence problem with very good results can be obtained with this method and [**?**] show the same outcomes of CMSA for the unbalanced common string partition problem. [**?**] apply CMSA to happy colourings and show that this approach can find good solutions on hard problem instances where exact approaches struggle. [**?**] develop a parallel hyrbid of CMSA and ACO and show that it is effective on project scheduling with the aim of maximising the NPV.

Merge Search was developed independently but uses similar ideas as that of CMSA at a high level. The crucial differences are in its ability to deal with extremely large populations of solutions and parallelisation. Like CMSA, the search iterative builds improving solutions to a problem by using the following steps: (a) maintain a population of feasible solutions, initially found through a heuristic (b) determine active variables in the MIP model from the population of solutions, (c) solve the resulting restricted MIP, thereby merging solutions, (d) use the solution information from the MIP to update the population, and (e) continue this process until some termination criteria is satisfied. Merge Search can be thought of as a generic matheuristic, combining integer programming and heuristic search, but in this paper we only focus on its application to the PCPSP (Precedence Constrained Pit Scheduling Problem).

This document is organised as follows. Section **??** discusses the details of the problem. This includes MIP formulations considered in this study and equivalences to others published in the literature. Next we introduce our new Merge Search heuristic and show how this can be applied to our problem in Section II. The details of how to apply this method to the PCPSP are described in Section **??**, including preprocessing, parallelisation and other implementational details that are important in obtaining good performance. Finally, we provide detailed computational results in Section **??** and show that we are able to produce both better feasible solutions and tighter bounds for most of the benchmarks that are available in the literature.

II. Merge Search

For extremely large problems such as open pit mining, a method that is able to efficiently combine solutions (potentially obtained from different sources) can be beneficial. This is particularly important in a parallel or distributed framework where multiple threads or processes are independently finding improved solutions. In such cases we do not want to just take the best solution, thereby discarding any improvements made by all other processes, but to learn from each of them. For this purpose, the key procedure proposed in this paper and used throughout the distributed solver method is the *Merge Search*. The aim of this is to combine the best features of multiple solutions that may have been arrived at independently, often by starting from the previously best known solution. Merge Search achieves this by solving an integer program over the space of combinations of solutions. That is it breaks all of the solutions into fragments (a set of variables) and re-assembles a new solution from these fragments.

Algorithm 1 Merge Search Matheuristic

Require: A combinatorial optimisation problem with variables x: $\max f(x) : x \in \mathcal{F}$
Require: Initial solution $x^0 \in \mathcal{F}$
for $k = 1, 2, \ldots$ **do**
  for $j = 1, 2, \ldots, m$ **do**
    Generate a solution s^j in a large neighbourhood of x^{k-1}
  end for
  Let $S = \{s^1, \ldots, s^m\} \cup \{x^{k-1}\}$ be all such solutions
  Let $\mathcal{P} = \{P_1, \ldots, P_p\}$ be a partition of the variables into sets for which all solutions are constant:

$$\bigcup_{P \in \mathcal{P}} P = \{1, \ldots, n\}, \qquad P \cap Q = \emptyset \; \forall \, P \neq Q \in \mathcal{P}, \qquad \text{and} \qquad s_i = s_j$$

  if $|\mathcal{P}| < K$ **then** split subsets until $|\mathcal{P}| = K$
  Solve $x^k = \arg\max f(x) : x \in \mathcal{F} \;\&\; x_i = x_j \; \forall i, j \in P \in \mathcal{P}$
end for

A. Description

Merge Search is a general matheuristic that operates as outlined in Algorithm 1. It carries out the following steps:

Finding an initial solution: In order to produce a population of solutions, an initial solution x^0 must first be produced. Some problems, such as the PCPSP considered here, are so large and complex, that even producing a feasible solution is quite computationally expensive — let alone one that is of guaranteed good quality. In contrast, other problems are structured such that producing a reasonable quality solution is not computationally expensive at all; however, often the methods of producing solutions to these problems are deterministic, creating issues with solution diversity. Although there are cases where finding a feasible solution is reasonably easy; in general, finding a feasible solution to a combinatorial problem is as hard as finding an optimal one [**?**].

When considering Merge Search as a general framework for solving constrained optimisation problems, the method by which the initial solution is produced is not important. The ideal circumstance would be when there already exists a custom heuristic for constructing a solution to the problem. However, if no such heuristic exists, then because it can be shown (Lemma 2) that the random splitting aspect of Merge Search makes it theoretically capable of producing any solution in the search space, it is possible to start with a completely random (feasible) solution and still find the optimal solution — if given enough time.

Of course, in practice, "enough time" can be infeasibly long for large search spaces, so a more intelligent method of producing an initial solution is required. Equation 1 gives a method of finding a feasible solution for problems which can be formulated as a mixed integer program (MIP) with n

variables and m linear constraints.

$$\begin{aligned} \text{minimise} \quad & \boldsymbol{c}^T\boldsymbol{x} + C^{max}\sum_{v_i \in \boldsymbol{v}} v_i \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b} - M\boldsymbol{v} \\ & \boldsymbol{x} \in \mathbb{Z}_2^n, \quad \boldsymbol{v} \in \mathbb{Z}_2^m. \end{aligned} \tag{1}$$

Here, $\boldsymbol{v} \in \mathbb{Z}_2^m$ is a vector and M is a constant sufficiently big enough such that $v_i = 1$ if constraint i is violated and $C^{max} = \sum_{c_i \in \boldsymbol{c}} |c_i|$ is a penalising constant that is added to the objective value every time a constraint is violated. The value of C^{max} is calculated by summing the magnitude of every value in the $\boldsymbol{c}$ vector, which means that it can be guaranteed that even the worst feasible solution will have a better objective value than the best infeasible one.

By solving this MIP model, a feasible solution to the problem is produced, which can then be used as an initial solution to generate a population[1].

Neighbourhood search (Step 3): Many solution merging meta-heuristics such as the construct, merge, solve and adapt (CMSA) heuristic [**?**] require a method that constructs solutions from scratch in order to produce a population to merge. However, for some large and complex problems, producing a feasible solution from scratch can be very computationally expensive. Merge search generates its population by defining a local-search operator for the problem and sampling the neighbourhood of a given solution; because it is generally easier to produce a new solution from an existing one, than it is to generate a solution from scratch.

The local-search neighbourhood can be as simple, or as sophisticated, as is required. It can be a custom-built, problem specific, heuristic that always produces feasible solutions; or it can be a heuristic that simply generates random bit-strings. So long as there is at least one feasible solution in the population, it can be shown (Lemma 1) that the merge operation will always produce a feasible solution and the random splitting heuristic ensures that any possible solution in the search space can be produced.

When considering Merge Search as a general meta-heuristic framework, it does not matter how the population is produced; some methods will produce populations that lead to more efficient searches, and some methods will produce populations that require a very long time to find the optimal solution. As with most hybrid meta-heuristic search techniques, it becomes about finding a good balance between how much computational effort is spent on exploration through, ensuring diversity of the population, and how much is spent on exploitation, through focusing on one particular area of the search space.

A generic neighbourhood search method can be defined for MIPs by fixing all but u variables (where u is a given fraction of the full set of decision variables) to the value of a given solution and solving this subproblem over the remaining u variables. Alternatively, a neighbourhood search can be performed generically using the machinery of Merge Search itself: define a random partition of variables, split that partition randomly and solve the subproblem induced by that partition. The number of neighbouring solutions m is an algorithm parameter.

Once a population has been generated, it can be used to define a partition on the decision variables.

Defining the partition (Step 6): The simplest way to partition the decision variables for a problem with a population of solutions $S = \{s_1, \ldots, s_m\} \cup \{x^{k-1}\}$ is to divide them into three groups: variables that *always* take the value 0 across all solutions; variables that *always* take the value 1 across all solutions; and variables that take *either* 0 or 1 across all solutions. This partition can be used to produce a reduced subproblem with the variables that are in the first group fixed to 0; the variables that are in the second group fixed to 1; and the variables that are in the third group allowed to take either 0 or 1.

In this sense, Merge Search can be thought of as a generalisation of an optimised, multi-parent uniform crossover operator similar to that used in the genetic algorithm (GA). In the uniform crossover operator for GA [**?**], each decision variable is considered in turn and its value selected from one of two parents with some probability. Because the value for every decision variable must be taken from either parent, no matter how many offspring are produced from these two parents the set of decision variables that are 1 for both parents will *always* take the value 1, and the set of variables that are 0 for both parents will *always* take the value 0. Effectively, these variables have been fixed and the only variables that are free to take either 0 or 1 are those that are not in either of these sets. This is exactly equivalent to performing a simple merge operation, without grouping, on a population consisting of two solutions; but where uniform crossover merely randomly samples the sub-space of solutions produced by the two parents, the merge operation searches that subspace for the (locally) optimal offspring. Uniform crossover is also *restricted* to a population of size two, whereas Merge Search generalises this idea to any arbitrary population size.

Using this naïve method of partitioning, with a very diverse population, when considering each variable in the $0/1$ region individually can result in very little reduction in the size of the subproblem produced. For this reason, a more sophisticated way of partitioning the decision variables is defined.

Definition 1. *Let x be the set of decision variables for a given problem, then the set $s_i \subseteq x$ is the set of decision variables that take the value 1 in a given solution i and $S = \{s_1, \ldots, s_m\}$ is a population of m solutions. Now, a* **merge partition** *is the set $\mathcal{P} = \{P \mid \bigcap_{i=1}^{m} s_i^{b_i},\ b \in \mathbb{Z}_2^m\}$ where $s_i^1 = s_i$ and $s_i^0 = s \setminus s_i$.*

Figure 1 illustrates all possible partitions that can be induced by a population of three solutions. The shaded circles indicate all of the variable assignments that are included in a particular solution. Here, partition $P_1 = s_1^0 \cap s_2^0 \cap s_3^0$, which corresponds to the set of decision variables that take the value 0 across

[1]Actually, for sufficiently large C^{max}, solving (1) is equivalent to solving the original problem (and will produce the same optimal solution) so, theoretically, there is no need to run the algorithm twice. In practice however, solving it in this way would be intractable for any problem that is large enough to be of interest, and so would not be a practical way of finding a feasible, initial solution. Any binary $\boldsymbol{x}$ can be used to produce a feasible initial solution for (1) by simply setting $v_i = 1$ iff the corresponding ith original constraint of $\boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b}$ is violated.

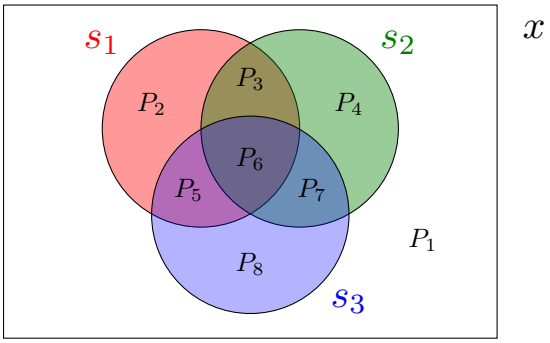


Fig. 1: Partitions formed by intersecting solutions.

all solutions; partition $P_6 = s_1^1 \cap s_2^1 \cap s_3^1$, which corresponds to the set of decision variables that take the value 1 across all solutions; and the other partitions correspond to sets of decision variables that, while they do not take the same value across all solutions, *agree amongst themselves* across all solutions. For example, all decision variables in partition P_5 take a 0 in s_2 and a 1 in s_1 and s_3.

By partitioning the decision variables in this manner, it is reasonable to assume that — given a big enough population size — were a new solution to be generated, the majority of the decision variables in the new solution would agree with the other variables in their respective partitions. Therefore, the decision variables can be aggregated into groups in the reduced subproblem, with each partition being considered as an individual variable.

The theoretical maximum number of partitions (and therefore, decision variables in the reduced subproblem) possible in a merge population of m solutions is 2^m. This theoretical maximum is only achieved when,

$$\forall (s_i, s_j) \in S^2,\ s_i^b \cap s_j^c \neq \emptyset, \forall b, c \in \mathbb{Z}_2;$$

however, typically not all solutions in a population will interact with all other solutions (i.e., there exist some pairs of solutions $(s_i, s_j) \in S^2$ such that $s_i^b \cap s_j^c \neq \emptyset, \forall b, c \in \mathbb{Z}_2$), so $|\mathcal{P}| << 2^{|S|}$, in practice.

Random splitting (Step 7): When generating a population by sampling the neighbourhood around an initial solution, the size of the partition induced by this set of solutions is typically quite small. As the size of this partition directly affects the size of the reduced subproblem, it is useful to be able to control the size of partition to increase the size of the merge neighbourhood that is searched. One way to do this is through a process called *random splitting*.

Definition 2. *Given a set $\mathcal{S}$, a* **random split** *is some heuristic process that produces two sets, $\mathcal{S}_1$ and $\mathcal{S}_2$, such that $\mathcal{S}_1 \cup \mathcal{S}_2 = \mathcal{S}$ and $\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$.*

This method of arbitrary splitting can be used to further partition the decision variables that have been aggregated, to allow the optimal solution to be produced. A proof of this is given in Lemma 2.

Simply splitting the partitions arbitrarily is unlikely to produce a useful partitioning, let alone the optimal one; therefore, it is beneficial to use some heuristic strategy to do it — this is especially true of large scale and highly constrained problems. For example, if x_a, x_b and x_c are decision variables in some partition P_i and there are constraints in the problem model that say $x_a \leq x_b \leq x_c$, it does not make any sense to split P_i such that $x_a, x_c \in P_i'$ and $x_b \in P_i''$, as the values of the merge variables representing P_i' and P_i'' in a merged solution would have to be equal in order to remain feasible. In this case, a random splitting heuristic should be designed that takes these precedence relationships into account.

There are many ways that a random splitting heuristic can be designed. The simplest is to generate a random bit string of length n and add it to the population before the partition is defined (Figure 2).

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
s_M	0	1	1	0	1	0	0	0	1	0	1	1	1	0	1	0
s_1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
s_2	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1
s_3	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1

Fig. 2: By adding the random bit string s_M to the population before merging, each partition is arbitrarily split into two.

In this figure, the "natural" partitions for the population $\{s_1, s_2, s_3\}$ are indicated by the blocks with shades of red, shades of blue and shades of green. By adding the random bit string s_M to the population, each natural partition has been arbitrarily split into two, indicated by the light and dark colour shades.

While this method is the simplest, it does not take into account any of the constraints, or the implicit structure of the problem. This means that, for highly constrained problems, the partitioning produced by this method is likely to be no more effective than the natural partitioning induced by the population itself. Therefore, in practice, it is wise to design a heuristic splitting method with these considerations in mind; however, if no such method is possible, it is theoretically possible to produce the optimal solution by simply using random splitting alone — when allowed enough time.

If defining a partition on a set of solutions can be said to be analgous to the crossover operator for GA, then splitting the partition is analogous to its mutation operator. As has been already established, crossover only allows solutions to be sampled from the sub-space induced by the properties of the two parents, not the entire search space; the same is true for defining a partition on the decision variables as described in the previous section — if $x_i = x_j$ across all solutions, any solution produced by merging in this way will also have $x_i = x_j$. In order to allow GA to produce any solution in the entire search space, a mutation operator is introduced; the simplest version of which selects a gene at random and flips its corresponding bit. This can be seen as a special case of the Merge Search heuristic, where a population consisting of a single solution and a single bit string with only one arbitrary

1 in it is merged. Here, the decision variables are partitioned into three groups: the group of variables that took the value 0 in the original solution; the group of variables that took the value 1 in the original solution; and the single variable to be "mutated". Again, as with the crossover analogy, whereas the mutation operator for GA merely samples the sub-space, Merge Search searches it to find the (locally) optimal choice.

By extending this idea further, it can also be shown that large neighbourhood search (LNS) [**?**], or indeed any *destroy-and-repair* heuristic, is a special case of Merge Search. To demonstrate this, a population is constructed that consists of a single solution s and a set of unique bit string masks $\{M_1, M_2, \ldots, M_m\}$, each with a single arbitrary bit flipped to 1, the rest of the bits are all zeroes.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
M_1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
M_2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
M_3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M_4	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
M_5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M_6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
s	1	0	0	1	1	0	0	0	1	0	0	1	0	0	1	1

Fig. 3: Each bit string mask separates its associated decision variable into a partition containing only that variable.

Figure 3 shows that when this population is merged, the decision variables are partitioned into $m+2$ groups: variables that took the value 0 in the original solution (red); variables that took the value 1 in the original solution (green); and m partitions containing a single variable, associated with each of the bit string masks (shades of blue). If the ratio of m to n is sufficient, such that enough of the original solution structure is maintained, then all those variables that took 0 or 1 in the original solution will be effectively fixed to their respective values in the reduced subproblem, and all those variables in the m singleton partitions are free to take either 0 or 1 in the solution to the reduced subproblem. This is equivalent to LNS, where a subset of the variables in a given solution is selected for "destruction" (i.e., removed from the solution) and the solution is "repaired" by solving the partial solution to (local) optimality.

Solution merging (Step 8): All of the steps leading up to this point have been working to construct a reduced subproblem, which can now be solved, using an exact solver or some other method, to produce a locally optimal solution for use in the next iteration of the process. In very general terms, the reduced subproblem takes the following form:

$$\begin{aligned} \text{minimise} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{F} \subseteq \mathbb{Z}_2^{|\boldsymbol{x}|}, \\ & x_i = x_j \quad \forall i, j \in P, P \in \mathcal{P}. \end{aligned} \tag{2}$$

Recall, S is the generated population of solutions and $\mathcal{P} = \{P_1, P_2, \ldots, P_p\}$ is the set of merge partitions produced by the population and the random splitting heuristic.

For problems with large numbers of decision variables, it can be practical to replace the set of problem variables x with a vector of partition variables z. This will often create a certain amount of overhead in constructing the model for the reduced subproblem as constraints need to be transformed to be in terms of partition variables instead of decision variables — and then again, when the produced solution must be re-expressed in terms of the problem variables. However, this usually results in the model taking up much less space in memory.

Finding the globally optimal solution to the reduced subproblem will give a locally optimal solution to the master problem Figure 4 gives an illustration of this process.

The partition on the decision variables representing the optimal solution $x^* \in \mathcal{F}$ is shown in blue in Figure 4a. Figure 4b shows the merge partitions produced by merging the population S and applying the random splitting heuristic. Here, the initial solution x^{k-1} is the straight red partition boundary and each neighbouring solution in the population is represented by two triangular regions protruding from either side of this boundary. These regions represent the decision variables that take different values in the neighbouring solution, with respect to the initial — regions on the left of the boundary indicate variables that have changed from 0 to 1 in the new solution, regions on the right indicate those that have changed from 1 to 0. The population is merged and split to produce the reduced subproblem which is expressed in terms of a set of partition variables z. This reduced subproblem is then solved using an exact solver or some other method. Figure 4c shows that the optimal solution to the subproblem z^* is a partition constructed from a subset of the boundaries of the merge partitions. Finally, z^* is mapped back to a solution to the master problem x^k (Figure 4c), an approximation of x^*. If the stopping criteria is not yet met, then the merged solution x^k is used as the initial solution for the next iteration.

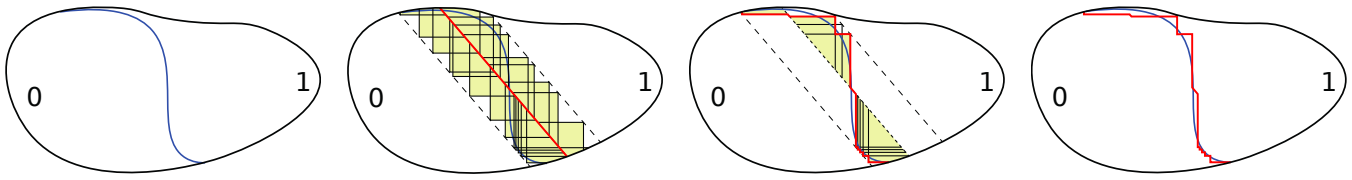
Provided there is at least one feasible solution in the population at the time of merging, the merged solution x^k will always be feasible. As the merge partitions are formed by the intersections of all the members in the population S (and then randomly split), any solution $s_i \in S$ can be reconstructed by simply including all partitions that are associated with that solution. This is illustrated in Figure 1 above, where s_1 can be constructed by the union of partitions $P_2 \cup P_3 \cup P_5 \cup P_6$. If s_i is feasible, then the final merge operation is able to produce s_i as x^k; meaning that, so long as the initial solution x^{k-1} is feasible, x^{k-1} will be a lower bound on the solution produced by merging and $f(x^{k-1}) \leq f(x^k) \leq f(x^*)$, even if the entire rest of the population consists of randomly generated bit strings, representing infeasible solutions.

B. Properties of Merge Search

There are some simple but important properties that follow directly from the way this matheuristic has been defined.

Lemma 1. *The Merge Step 8 produces a solution that is at least as good as any of the neighbours in S: $f(x^k) \geq f(s)\ \forall s \in S$*

Proof. By construction any solution $s \in S$ satisfies $s \in \mathcal{F}$ and $s_i = s_j\ \forall i, j \in P \in \mathcal{P}$ and so is feasible for the optimisation problem in Step 8. Hence, $f(x^k) \geq f(s)$. □



(a) Optimal solution x^* (b) Merge partitions $\mathcal{P}$ (c) Sub-problem solution z^* (d) Merged $f(x^k) \leq f(x^*)$

Fig. 4: Solving the reduced subproblem induced by the merge partitions $\mathcal{P}$ to optimality, gives an approximate solution to the master problem.

Corollary 1. *Merge Search is a hill-climbing method that produces a non-decreasing sequence of solution values:* $f(x^0) \leq f(x^1) \leq f(x^2) \leq \ldots$

Hence, for diversification, the method relies entirely on the neighbourhood search in Step 3 of the algorithm and the randomised splitting in Step 7. This is not a problem for large instances such as this, where simply finding a very good local minimum is already a very challenging task. Furthermore, the following property holds:

Lemma 2. *For pure binary problems where all variables are* $0-1$*, if random splitting of subsets is used in Step 7 (allowing any possible split with some non-zero probability) with* $K \geq 2^{m+2}$ *then Algorithm 1 is guaranteed to converge to the optimal solution as the number of iterations approaches infinity.*

Proof. By assumption there exists an optimal solution in $\mathcal{F}$, due to the existence of $x^0 \in \mathcal{F}$ and boundedness of $\mathcal{F}$ when all variables are binary. Let s^* be any optimal solution. The first thing to note is that if a partition $\mathcal{P}^*$ satisfies $s^*_i = s^*_j \; \forall P \in \mathcal{P}^*, \; i,j \in P$ then the Merge problem in Step 8 yields and optimal solution. Such a partition could be generated, from any arbitrary partition $\mathcal{P}$ by splitting each $P \in \mathcal{P}$ into $P_0 = P \cap \{i \mid s^*_i = 0\}$ and $P_1 = P \cap \{i \mid s^*_i = 1\}$. This splitting at most doubles the number of elements in the partition.
Now the same argument can be used to show that $|\mathcal{P}|$ as produced in Step 6 has at most 2^{m+1} elements (corresponding to splitting based on solutions $s^1, \ldots, s^m$ and x^{k-1}). Hence, we only need K to be at least 2^{m+2} to allow some chance of generating the required partition in any step. Hence, as the number of iterations of Algorithm 1 goes to infinity the chance of *not* producing an optimal solution goes to zero. And of course, based on Corollary 1 once an optimal solution has been found, the algorithm will not depart from this. □

While convergence to the optimal solution is of course extremely unlikely for practical sized instances, the lemma shows that it is at least theoretically possible. Furthermore, while 2^{m+2} appears quite large, the only real requirement is that K is sufficiently large to allow each element of $\mathcal{P}$ to be split once, with $|\mathcal{P}|$ typically much smaller than 2^{m+1} in practice. Hence, while Merge Search is a hill-climbing method, it is at least in principle possible for the method to reach a global optimum from any starting point.

C. Comparison with CMSA

As mentioned previously, it may be noted here that this meta-heuristic has some similarity to the recently published CMSA heuristic by [**?**]. Starting with an empty subinstance of the problem $\mathcal{C}'$, solutions are probabilistically generated in this method, from scratch, and their components added to $\mathcal{C}'$. This reduced subinstance is then solved using an exact solver and a so-called "ageing" mechanism is used to remove components from $\mathcal{C}'$ that have not been useful in the preceding iterations.

Although there are some similarities between Merge Search and CMSA, there are two areas where Merge Search diverges significantly from it. These are:

- the generation of candidate solutions to be merged; and
- the aggregation of decision variables in the reduced subproblem.

The CMSA subproblems are still defined based on a population of solutions, however these solutions are constructed probabilistically, with each solution being generated from scratch. This has the two-fold effect of reducing the capacity of CMSA to learn from the best individual solution found so far, and also makes it impractical when trying to solve large-scale, or very complicated, problems for which constructing a feasible solution is extremely time consuming, such as the PCPSP that is considered here. Merge search avoids this issue by heuristically constructing an initial solution and then sampling its neighbourhood in order to generate a population, which is then used to define its subproblems. The effect of this is to make the time taken to generate the population of solutions dependent on the local search algorithm used to sample the neighbourhood of a given feasible solution which is often much faster than the algorithm used to find feasible solutions from scratch. Of course, this leaves Merge Search potentially susceptible to being overly sensitive to the quality of the initial solution — as discussed previously in this section — however, this can be mitigated by increasing the diversity of the population, or further splitting the merge partitions.

The second area where there is a significant divergence between the two methods is in the aggregation of decision variables. CMSA uses its population of constructed solutions to determine the variables that are to be included in its reduced subproblem. If a variable is represented by an element of one of the candidate solutions in the population, it is automatically included in the reduced subproblem. These variables are added to the subproblem individually, and as such this aspect functions similarly to large neighbourhood search (LNS). One consequence of this is that the subproblems can become very

large, especially for problems where there are naturally many non-zero values in a solution. To combat this, a so-called "ageing" mechanism is introduced in CMSA to ensure that elements that have not been useful in producing good quality solutions recently are removed from the pool of solution elements. In contrast, Merge Search uses information from across the entire population to determine which variables are added into the reduced subproblem as well as to aggregate variables that share common values across the population, and so likely share some kind of dependency. This grouping allows for more compact subproblems and a much larger region of the search space can be covered for the same computational power. The trade-off for this is much coarser-grained solutions, however this can be alleviated by the introduction of random splitting to these groups to help escape local optima.

III. Applying Merge Search to solve problems

Merge Search is presented here as a general hybrid meta-heuristic framework for solving constrained combinatorial optimisation problems. In order to demonstrate its effectiveness and versatility, Merge Search is applied to two problems from two different domains: the constrained pit-limit (CPIT) problem, an abstraction of a real-life problem from open-pit mining; and the Steiner tree problem in graphs (STPG), a famous NP-Complete problem which is very well-studied in the literature.

A. Open-pit mining problems

Open-pit mining is a very important industry in Australia and around the world [**?**]. Two of the most critical tasks within the life-cycle of an open-pit mine is planning and production scheduling. These tasks allow the mine operator to estimate the total value of the mine over its life and also to identify areas for excavation that will yield the most value. Proper planning of a mine ensures maximum profit for the operator and, because this is typically talked about in the hundreds of millions of dollars, it is an excellent application for optimisation techniques as very small changes in efficiency can still translate to significant sums of money.

In order to model something so complex as a combinatorial optimisation problem, the earth to be mined (known as the *orebody*) is typically discretised into a three-dimensional array of *blocks* with each assigned a value based on the ore content and the cost required to excavate it. These values are calculated by taking core samples and using geological and statistical methods to estimate the value of each block. The aim is to maximise the net present value (NPV) of the mine by determining the set of blocks to extract and the order in which to extract them [**?**].

Problems in mine planning and production scheduling are very large and tend to have few side-constraints (often well under a hundred), but many blocks and many, many more precedence constraints governing when blocks can be mined. This means traditional mathematical solvers are unable to solve these problems without first using some form of decomposition, making these problems perfect candidates for hybrid meta-heuristics, despite there being very little in the literature.

Due to the sensitive nature of information surrounding mining enterprises, obtaining problem data for academic research can be challenging, however the website *minelib* [**?**] has a repository of problems and results that are freely available to the general public. These sets contain data for versions of the problem such as the ultimate pit limit (UPIT), constrained pit-limit and the precedence constrained production scheduling problem (PCPSP). It is the CPIT problem that will be the focus of these experiments.

1) Solving CPIT with Merge Search: [**?**] describe a novel representation for the this type of problem and a greedy randomised adaptive search procedure (GRASP) algorithm for solving it. They build on this with local search operator for the CPIT problem, and use it in a simple merge search algorithm that operates without variable partitioning or random splitting [**?**]. By incorporating a variable partitioning mechanism, a much larger region of the search space is able to be explored for the same computational budget [**?**]. They exploit the structure of the problem, and treat many decision variables as a single group. This allows the size of the mixed-integer programming sub-problem to be greatly reduced and hence, the time required to solve it.

The algorithm presented in this paper is an exentsion of this work, with two additions: a random splitting heuristic, to allow greater granularity in the solutions produced; and a "solution polishing" technique, based on the local improvement heuristic from the GRASP algorithm.

B. The Steiner tree problem in graphs

The Steiner tree problem in graphs (STPG) is a classic problem in the field of combinatorial optimisation, the decision version being one of the original 21 NP-complete problems outlined by [**?**] in his seminal paper. Aside from being a fundamental problem in the abstract world of computing theory, the STPG has numerous real-world applications in communications, pipeline and transport network design, computational biology and very large-scale integrated circuit (VLSI) design [**?**].

Despite being around for centuries [**?**], the STPG has also gained much attention in recent decades due to it being the mathematical structure behind multicast networking problems [**?**]. Exact methods for solving the STPG have been developed using techniques such as integer linear programming (ILP), lagrangian relaxation and primal-dual strategies [**?**]; however these approaches suffer from exponential worst-case computation times which can make some large-scale instances intractable. The current state-of-the-art exact approaches to solving the STPG are hybrid [**?**], [**?**]; several algorithmic, graph reduction, metaheuristic and mathematical programming techniques, working together to produce provably optimal solutions in a much faster time than traditional optimisation techniques alone. When good quality, but not necessarily optimal, solutions are required in a reasonable amount of time, metaheuristics and decomposition techniques have been used to tackle this problem; some of the more successful approaches in this manner have been memetic algorithms [**?**], ant colony optimisation [**?**], [**?**], local search techniques [**?**], [**?**] and voronoi-based decomposition heuristics [**?**].

The STPG is extremely well-travelled, with many sophisticated pre-processing techniques and methods of finding good quality solutions already available in the literature. The purpose of including it in this paper is not to improve the current state-of-the-art results; but to provide a simple test-bed, upon which the properties of the proposed merge search algorithm can be investigated — much more easily than with a more complex, real-world problem such as the constrained pit-limit problem, mentioned above.

1) Solving the STPG with Merge Search: First described by [**?**], but subsequently used widely by many researchers, is the so-called key path neighbourhood. A key path is defined as a path within a Steiner tree where the two end vertices are either terminal vertices or vertices of degree at least 3; all intermediate vertices (if any) are of degree 2 and are not terminal. The useful property of such structures is that their removal from a solution to the STPG will always result in two disconnected trees which can then be subsequently reconnected. This local search neighbourhood was extended by [**?**], by adding a so-called "jump" operator which aids in escaping local minima, and is used as the basis of the Merge Search algorithm presented in this paper.

The problem instance is first pre-processed using some of the techniques described in [**?**], [**?**], [**?**] to reduce the number of decision variables. An initial solution is constructed and a neighbouring population produced, using the methods described in [**?**]. This population of solutions is used to partition the decision variables in the manner described in Section II.

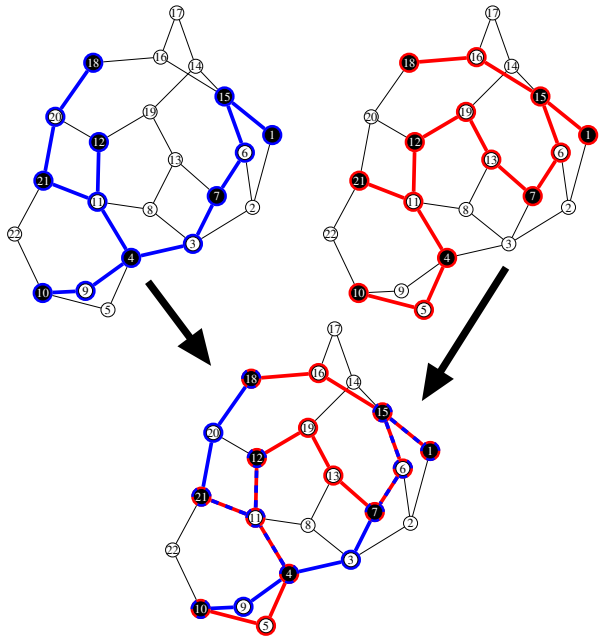


Fig. 5: By merging two solutions (top left and right), the decision variables are partitioned into four disjoint sets, indicated here (bottom) by: red; blue; red and blue; and no colour.

These partitions can be split by selecting an arbitrary partition, and an arbitrary variable in that partition, and separating all variables representing vertices and edges that are connected to that arbitrary variable. This ensures that any partition is split such that all variables removed from the original partition comprise a connected sub-graph, and as such should be able to be included in the resultant merged solution on its own, without the original partition. If a partition is split such that all variables are separated, leaving the original partition empty, then the last variable added to the split is kept in the original partition. As the connected variables are added using DFS, this is always a variable on the boundary, and so this method will not suffer from the problems present with truely random splitting.

Having partitioned and split the decision variables, the resulting reduced subproblem can be solved using a version of the [**?**] MIP formulation that has been modified be expressed in terms of a single set of partition variables, instead of edge and vertex variables.

IV. Experiments

This section details the experimental set up, the experiments performed and finally presents the results and discusses the implications those results suggest.

A. Datasets

The CPIT problem:: Table I details the properties of the problem instances used to test the algorithm in this paper. The instance name is given in the first column; followed by the number of blocks in the orebody model; the number of precedence arcs for each instance is provided in the third column; the fourth column gives the total number of time periods available; the number of decision variables is shown in the second last column; with the last column giving the total number of constraints in the problem model.

TABLE I: Characteristics of *minelib* [**?**] datasets.

Instance	Blocks	Precedences	Periods	Variables	Constraints
newman1	1,060	3,922	6	6,360	29,904
zuck_small	9,400	145,640	20	188,000	3,100,840
kd	14,153	219,778	12	169,836	2,807,196
zuck_medium	29,277	1,271,207	15	439,155	19,507,290
marvin	53,271	650,631	20	1,065,420	14,078,080
zuck_large	96,821	1,053,105	30	2,904,630	34,497,840

A few instances from the full *minelib* dataset were omitted due to missing files, referencing more than two resources or being too large for the algorithm in its current incarnation.

The Steiner tree problem in graphs:: The experiments were were carried out on categories B, C, D and E of the STPG problems from the *steinlib* library [**?**], a standard dataset used by many researchers. The smaller-scale, category B instances are 18 randomised networks with 50 to 100 vertices and 63 to 200 edges. The C and D datasets consists of 20 larger randomised networks, with 500 and 1,000 vertices, respectively, and between 625 to 25,000 edges. Finally, the E dataset contains the largest instances, each with 2,500 vertices and between 3,125 and 62,500 edges. Details of the individual instances can be found in Table II.

The optimal solutions for these instances are reported in the library and have been proven by exact methods such as branch-and-bound with graph reduction techniques.

TABLE II: Characteristics of category B and C instances from the *steinlib* database. $|V|$: number of vertices; $|E|$: number of edges; $|T|$: the number of terminals; and, opt: the total cost of the optimal solution.

Category B					Category C					Category D					Category E				
Inst.	$\|V\|$	$\|E\|$	$\|T\|$	opt	Inst.	$\|V\|$	$\|E\|$	$\|T\|$	opt	Inst.	$\|V\|$	$\|E\|$	$\|T\|$	opt	Inst.	$\|V\|$	$\|E\|$	$\|T\|$	opt
$b01$	50	63	9	82	$c01$	500	625	5	85	$d01$	1,000	1,250	5	106	$e01$	2,500	3,125	5	111
$b02$	50	63	13	83	$c02$	500	625	10	144	$d02$	1,000	1,250	10	220	$e02$	2,500	3,125	10	214
$b03$	50	63	25	138	$c03$	500	625	83	754	$d03$	1,000	1,250	167	1,565	$e03$	2,500	3,125	417	4,013
$b04$	50	100	9	59	$c04$	500	625	125	1,079	$d04$	1,000	1,250	250	1,935	$e04$	2,500	3,125	625	5,101
$b05$	50	100	13	61	$c05$	500	625	250	1,579	$d05$	1,000	1,250	500	3,250	$e05$	2,500	3,125	1,250	8,128
$b06$	50	100	25	122	$c06$	500	1,000	5	55	$d06$	1,000	2,000	5	67	$e06$	2,500	5,000	5	73
$b07$	75	94	13	111	$c07$	500	1,000	10	102	$d07$	1,000	2,000	10	103	$e07$	2,500	5,000	10	145
$b08$	75	94	19	104	$c08$	500	1,000	83	509	$d08$	1,000	2,000	167	1,072	$e08$	2,500	5,000	417	2,640
$b09$	75	94	38	220	$c09$	500	1,000	125	707	$d09$	1,000	2,000	250	1,448	$e09$	2,500	5,000	625	3,604
$b10$	75	150	13	86	$c10$	500	1,000	250	1,093	$d10$	1,000	2,000	500	2,110	$e10$	2,500	5,000	1,250	5,600
$b11$	75	150	19	88	$c11$	500	2,500	5	32	$d11$	1,000	5,000	5	29	$e11$	2,500	12,500	5	34
$b12$	75	150	38	174	$c12$	500	2,500	10	46	$d12$	1,000	5,000	10	42	$e12$	2,500	12,500	10	67
$b13$	100	125	17	165	$c13$	500	2,500	83	258	$d13$	1,000	5,000	167	500	$e13$	2,500	12,500	417	1,280
$b14$	100	125	25	235	$c14$	500	2,500	125	323	$d14$	1,000	5,000	250	667	$e14$	2,500	12,500	625	1,732
$b15$	100	125	50	318	$c15$	500	2,500	250	556	$d15$	1,000	5,000	500	1,116	$e15$	2,500	12,500	1,250	2,784
$b16$	100	200	17	127	$c16$	500	12,500	5	11	$d16$	1,000	25,000	5	13	$e16$	2,500	62,500	5	15
$b17$	100	200	25	131	$c17$	500	12,500	10	18	$d17$	1,000	25,000	10	23	$e17$	2,500	62,500	10	25
$b18$	100	200	50	218	$c18$	500	12,500	83	113	$d18$	1,000	25,000	167	223	$e18$	2,500	62,500	417	564
					$c19$	500	12,500	125	146	$d19$	1,000	25,000	250	310	$e19$	2,500	62,500	625	758
					$c20$	500	12,500	250	267	$d20$	1,000	25,000	500	537	$e20$	2,500	62,500	1,250	1,342

B. Pre-processing

The CPIT problem: The "time-expanded" method of representing the CPIT problem given in [**?**] has one significant drawback: it exponentially increases the size of the problem instances. In this technique, each block-time pair is associated with a single decision variable and a solution is represented as a closure which partitions this expanded graph. In order to mitigate this effect, the structure of the problem can be exploited to apply pre-processing to reduce the number of variables and constraints in each problem instance.

First, a relaxed version of the problem, called UPIT, is solved to determine the set of blocks that are worth extracting at all; then the earliest and latest possible times for extraction for each block are computed. Because a condition of extracting a block is that all blocks above it must be extracted first, blocks that are deep cannot be extracted too early in the process — because there is not enough time to reach them — and similarly, blocks that are close to the surface cannot be extracted too late — as there will not be enough time to reach the blocks below. More detail about these pre-procesing methods can be found in [**?**].

The Steiner tree problem in graphs: There are many techniques available in the literature for pre-processing instances of the STPG, ranging from the very simple to the very complicated. As the experiments on the STPG for this study were not intended to improve the current state-of-the-art, but merely to investigate the properties of merge search and prove its versatility; the purpose of pre-processing the problem instances was simply to make them more manageable for the algorithms being tested. To this end, three basic ones were chosen from the literature: removing non-terminal vertices of degree 1; removing edges where a shorter path exists in the graph; and, replacing the edges incident to non-terminal vertices of degree two with a single edge. For further reading on graph reduction techniques for pre-processing instances of the STPG, see [**?**], [**?**], [**?**].

C. Experimental setup

The experiments were carried out on an *Intel® Core™ i5-2320* processor (3.0GHz) with 24GB RAM running Linux. All code was implemented in C++ with GCC-4.8.0. CPLEX Studio 12.7, operating with a single thread due to the need for callbacks, was employed as the mixed integer programming (MIP) solver. For the CPIT problem experiments, the boost library implementation of the Boykov-Kolmogorov algorithm was used to solve the UPIT sub-problem during the pre-processing stage.

The CPIT problem:: The merge search algorithm was compared against the baselines of the results published on the *minelib* [**?**] website and the results from using a greedy randomised adaptive search procedure (GRASP) heuristic for the precedence constrained production scheduling problem (PCPSP) [**?**], adapted for use with the CPIT problem. Also provided are the results for a variant of the merge search algorithm that uses variable grouping but no random splitting, previously published in [**?**] — these results are included in order to illustrate the effect that random splitting has on solution quality.

Each algorithm was run 20 times on each instance, recording the mean objective value and standard deviation of the resulting solution produced by each run. Finally, the last experiment performed was to use the local improvement heuristic from the GRASP algorithm to "polish" the best result produced by the merge algorithm.

The Steiner tree problem in graphs:: The merge search algorithm was compared against three separate baseline algorithms, the greedy randomised adaptive search (GRASP) heuristic, pure local search and pure MIP. The method of constructing initial solutions was different between the GRASP and merge search algorithms, so pure local search and pure MIP were included to ensure any improvement was not solely based on this factor. Aside from population merging, the merge search algorithm comprises two main components, local search and MIP search; so by isolating these two factors, it can be

shown that merge search is greater than the sum of its parts.

Each algorithm was run 30 times on each (pre-processed) instance from the datasets, recording the mean objective value and standard deviation across all of the runs. All instances in the datasets used are supplied with their optimal objective values and this is used for the main stopping criteria. Otherwise, the stopping criteria of the merge search algorithm is generally dictated by the number of seconds spent in the MIP search so the time taken for each search is not provided, as it does not give much information about the performance of the algorithm. However, special mention is made when the algorithm terminated early for all runs.

Additional experiments were performed to investigate the difference between using the random and deterministic solution construction heuristics and the effect of population size on the size of the reduced sub-problem. In order to preserve space, a representative subset of the problem instances is used to illustrate the outcome of these experiments; however, these results are not "cherry-picked" and the full tables are available in the appendix.

V. Results and discussion

A. The CPIT problem:

Table III provides the results of the three algorithms tested on the six problem instances from the *minelib* dataset, along with the linear programming (LP) upper bound and the current best solution as published on the *minelib* website, ordered from smallest problem instance to largest.

In this table, it can be seen that the two merge search algorithms consistently produce better results than the *minelib* and the GRASP heuristic results, except for `newman1` and `zuck_large`. The results for `newman1` are similar because the problem instance is so small and it can be assumed that the objective value of 2.418E+07 is quite close to the optimal solution, as all three algorithms agree on this as an average value and when the search is performed without any stopping criteria, the maximum objective value that is found is 2.41798E+07.

Comparing the results of the two variants of merge search illustrates the effect that random splitting has on the quality of the solution produced. For smaller problem instances, the two algorithms produce very similar quality solutions, indicating that the random splitting has little effect. However, as the size of the problem increases, the effect of the random splitting becomes more pronounced, with the biggest effect being on the two largest instances `zuck_medium` and `zuck_large`[2].

If there are no overlaps at all between variables across solutions, then the partitions in the reduced sub-problem are simply the set differences of the variables in each solution and the seed solution. In this extreme case, the merge operation is unable to produce a solution that does not already exist in the population, unless some random splitting of the partitions is performed. As the size of the problem instance decreases, the likelihood that there will be overlaps between variables across solutions increases. These overlaps in the variable sets produce splits in the partitions when producing the reduced sub-problem, reducing the need for additional random splitting to produce a different solution to those already existing in the generated population. This is not to suggest that random splitting would not be beneficial; but the problem with random splitting is that it is *random*, and there is no way of guaranteeing that a particular split will improve the quality of a solution after merging, any more than a split produced by overlapping variable sets will.

The GRASP heuristic produces consistently worse results than merge search. This is expected as the sliding window heuristic used in the local improvement phase of the algorithm is better suited to incrementally improving a good solution than turning a mediocre solution into a good one, as it only operates on a small subset of the variables at one time. This means that it is good for "tweaking" a solution by shifting the time that a block is mined forward or backward one or two periods; but it is no good if the time that the block is mined must be moved by many periods, as this would take a lot of passes to achieve.

Convergence behaviour: This idea of GRASP being better suited to incrementally improving an already good solution is demonstrated in Figure 6. This figure gives the plots of the convergence behaviour of both the GRASP algorithm and merge search on all six of the CPIT instances. It can be seen in these plots that merge search converges quicker in nearly all instances, except for `kd`.

The difference here, is that the initial solution is already quite close in value to the resultant solution, so there are not a lot of improvements that can be made. This means that the more exhaustive search for small improvements of the GRASP algorithm is more likely to be effective, early on, than the more global search of the merge search algorithm, which relies on the stochastic natures of the local search operator and arbitrary splitting heuristic to find its improvements. Merge search does get there in the end — and manages to find a slightly better solution in this case — but it takes longer and with a more gentle curve. For all the other instances, where the gap between the initial solution and the resultant one is much larger, merge search is demonstrably more suited to the task.

It can also be seen from these plots that once the algorithm has seemingly converged, it can sometimes find a way out of the local optima and find a much better solution. This is evidenced in the plots for `zuck_medium` (Figure 6d), `marvin` (Figure 6e) and also Figure 8a below. This behaviour is expected in such large problem instances, as once it has started to converge there are many ways to make the solution worse, but only a few to make it better; but once it has found a way out of the local optima, often several other improving moves will become apparent as well.

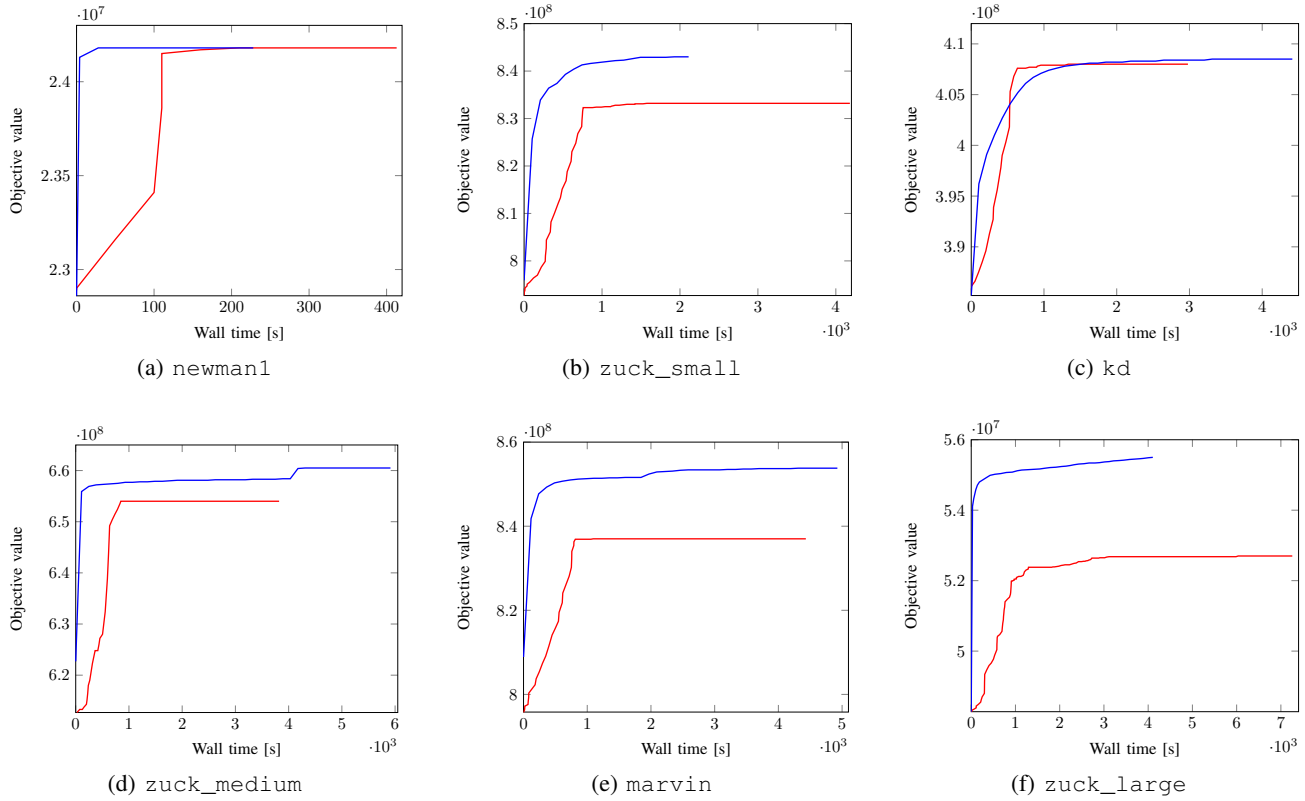
Runtime information: Table IV gives the runtime information (wall time and CPU time) for both the GRASP and merge search algorithms. As they use a MIP solver to solve their restricted sub-problems, the runtime of both GRASP and

[1]This result was beaten by polishing the best merge search solution with the GRASP local improvement heuristic (Table V).

[2]Although it is technically a larger problem than `zuck_medium`, solving the UPIT problem on the `marvin` instance, as part of the pre-processing stage, eliminates many blocks and makes its effective size much smaller than `zuck_medium`.

TABLE III: Results on *minelib* dataset instances. Given are the LP upper bound, current best solution from the *minelib* website, the mean objective values (μ) and standard deviations (σ) for the GRASP heuristic, merge search without random splitting and full merge search heuristic.

	LP UB	*minelib*	GRASP heuristic		merge search (no splitting)		merge search (with splitting)	
Instance			μ	σ	μ	σ		
newman1	2.449E+07	2.348E+07	**2.418E+07**	4.331E+03	**2.418E+07**	1.165E+02	**2.418E+07**	1.012E+03
zuck_small	8.542E+08	7.887E+08	8.334E+08	2.951E+06	**8.433E+08**	2.078E+06	8.432E+08	1.747E+06
kd	4.095E+08	3.969E+08	4.083E+08	6.135E+05	**4.088E+08**	1.881E+05	4.086E+08	4.516E+05
zuck_medium	7.106E+08	6.154E+08	6.548E+08	3.764E+06	6.584E+08	8.462E+05	**6.610E+08**	4.707E+06
marvin	8.639E+08	8.207E+08	8.368E+08	9.006E+06	8.539E+08	2.575E+05	**8.540E+08**	4.196E+05
zuck_large	5.739E+07	**5.678E+07***	5.269E+07	1.297E+06	5.441E+07	2.558E+05	5.524E+07	5.890E+05



(a) newman1 (b) zuck_small (c) kd

(d) zuck_medium (e) marvin (f) zuck_large

Fig. 6: Convergence plots for merge search on CPIT instances. Merge search data is shown in blue and GRASP is shown in red.

merge search is reasonably easily configured by controlling how long the solver is allowed to run for each iteration. The parameters of the GRASP algorithm were chosen so that the search would take roughly the same amount of time as that of the merge search — and for the most part, they are pretty similar.

The two instances that are significantly different in runtime between GRASP and merge search are kd and zuck_large. The reason that zuck_large is so different is that it was impossible to run the merge search algorithm with a population of 1,000 on such a large problem instance, due to memory issues (even with 23 GB of RAM!); so a smaller population size of 500 was used, which took much less time to produce and to compute the merge partitions.

The reason for the differences in runtime on kd is illustrated in Figure 7. This figure shows plots of the amount of wall time

TABLE IV: Runtime information for experiments on CPIT. Given are the mean (μ) and standard deviations (σ) of the wall and CPU time for the GRASP and merge search algorithms, in seconds.

	GRASP				merge search			
	Wall time [s]		CPU time [s]		Wall time [s]		CPU time [s]	
Instance	μ	σ	μ	σ	μ	σ	μ	σ
newman1	453.2	33.8	1,430.6	88.0	227.5	17.6	710.9	41.7
zuck_small	4,215.3	152.5	12,128.7	563.3	4,214.2	188.6	9,938.5	636.8
kd	3,023.2	307.5	8,459.5	844.9	4,378.3	175.3	11,525.9	845.0
zuck_medium	4,021.7	252.3	9,123.6	478.4	5,487.0	248.8	16,843.3	745.1
marvin	4,388.8	252.6	14,653.2	753.4	4,867.8	214.7	15,072.5	396.4
zuck_large	7,081.3	453.8	18,066.7	1,399.9	3,978.3*	152.1	9,745.8	786.9

*population size for zuck_large set at 500 for merge search due to insufficient memory.

in seconds consumed per iteration of the respective algorithms. The GRASP algorithm is measured by "window movements", which counts the number of times the window has incremented

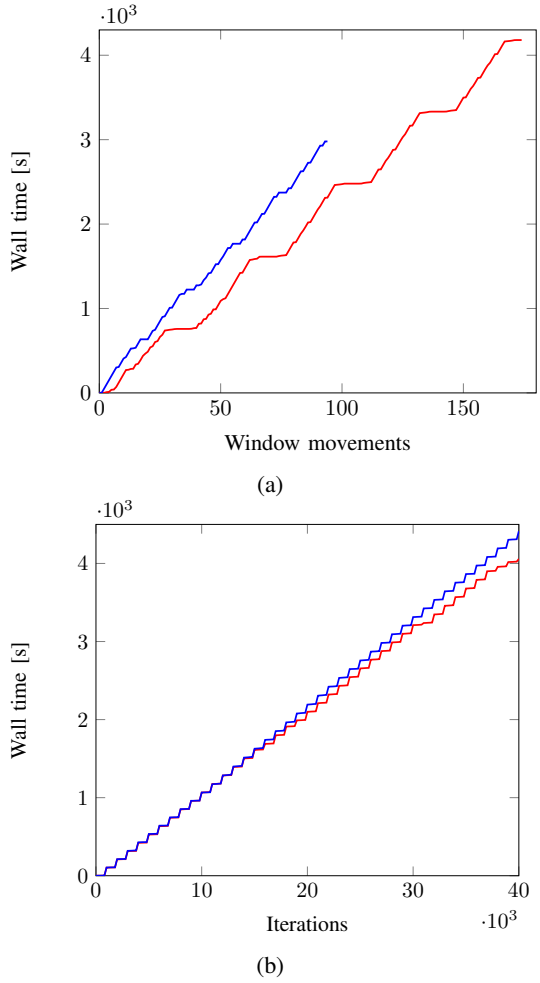


Fig. 7: Plot of wall time [s] for `zuck_small` (red) and `kd` (blue) instances. Sub-figure (a) shows GRASP algorithm data and (b) shows merge search data.

forwards or backwards along the solution. The merge search algorithm is measured by iterations; each the time for each iteration is measured at the point when a new candidate solution is generated for the population, so it is clear to see that, with a population size of one thousand, every thousandth iteration will include an extra amount of time to include the merge operation itself.

It is important to remember here that the `kd` instance only has a maximum of 12 periods allowed, while `zuck_small` has 20. So the difference in time between GRASP and merge search is easy to explain, because GRASP is dependent on the number of periods in the solution as that determines how many window movements will be made; whereas merge search will generate the same size population and perform the same number of merges, no matter how many periods are allowed. This is demonstrated very clearly in Figure 7b, where there is very little difference between the two instances. It can be seen that the population is generated reasonably quickly (indicated by the flat parts of the curve) and then each point where there is a sudden jump in the amount of time taken indicates that a merge operation has taken place.

But this is not the end of the story. In these two examples, GRASP took 2,979 seconds to complete the search on `kd` and it took 4,180 to complete the search on `zuck_small`. If there was a direct linear relationship between time taken to complete the search and number of periods, the amount of time GRASP should take to complete its search on `zuck_small` should be $\frac{2{,}979\times 20}{12} = 4{,}965$. So where did the other 13 minutes go?

Looking at Figure 7a, the main thing that can be noticed here is that the shape of the plot for `zuck_small` is significantly different to the shape of the plot for `kd`. The regular flat spots in the plot for `zuck_small` indicate that there is a group of periods for that problem that are not used by the solution, or do not contain any blocks that can be moved around, and therefore the MIP solver will not take the full amount of allowed time to solve the reduced sub-problem. It so happens that the solution to `zuck_small` does not use any period after period 16, so the first and last few window moves of each full pass take very little time at all. In contrast the `kd` instance uses periods right up until period 10, so there are fewer window moves that will be skipped over, as evidenced by the straighter line on the plot.

Solution polishing: As the quality of the solutions produced by the random construction heuristic is not extremely high, too much of the local improvement phase is spent moving the times that blocks are mined over long temporal distances; so the algorithm is unable to produce high quality solutions in the allowed computational budget. It was observed however, that the quality of the solutions produced by the GRASP algorithm greatly depended on the quality of the initial solution that was constructed. So it was decided to see what would happen if the local improvement heuristic from the GRASP algorithm was applied to "polish" the best solution obtained by the merge search algorithm. These results are given in Table V.

TABLE V: Results of polishing the best merge search solution with the local improvement heursistic from the GRASP algorithm. Given is the LP upper bound and current best solution from the *minelib* website, the objective value of the best solution produced by the merge search algorithm and the value of that solution when polished by the local improvement heuristic.

Instance	LP UB	*minelib*	merge search	polished merge
`newman1`	2.449E+07	2.348E+07	**2.418E+07**	**2.418E+07**
`zuck_small`	8.542E+08	7.887E+08	8.456E+08	**8.471E+08**
`kd`	4.095E+08	3.969E+08	**4.091E+08**	**4.091E+08**
`zuck_medium`	7.106E+08	6.154E+08	6.669E+08	**6.688E+08**
`marvin`	8.639E+08	8.207E+08	8.551E+08	**8.564E+08**
`zuck_large`	5.739E+07	5.678E+07	5.625E+07	**5.730E+07**

This table clearly demonstrates that the sliding window heuristic is very effective in improving solutions produced by merge search, even surpassing the *minelib* result for `zuck_large` — something that merge search could not achieve on its own, without polishing. For these experiments, the best solution from the merge search runs is taken and then three passes of the sliding window heuristic is applied. Although doing this effectively increases the allowed computational budget, the quality of the solutions produced by doing this is better than if the computational budget is increased for either of the two algorithms on their own. In the case of the GRASP algorithm, the quality of the solution at the end of its

run is still too low to be able to improve the objective value by very much in only three passes. In the case of the merge search algorithm, because both the population generation and merge partition splitting occurs stochastically, as the quality of the solution increases the chances of finding an improving solution, or partition split, by chance decreases. This means that the search is likely to have converged by the end of its run, so increasing the computational budget at this point is unlikely to produce much of an improvement in the quality of the solution, if any at all.

Applying the sliding window heuristic at the end of the merge search process solves both of these problems. The sliding window heuristic performs a much more methodical and exhaustive search to find small improvements that can be made to the solution which might be difficult to find by random sampling, in such a large search space. Additionally, by starting the sliding window heuristic with a much higher quality initial solution, computational effort does not need to be wasted on the "low hanging fruit" that can be found easily by sampling the search space. As the heuristic only considers a small subset of the variables at a time, it cannot make large changes to the solution in a single iteration.

Figure 8 contains plots that illustrate this concept on two instances of the CPIT problem, `zuck_medium` and `zuck_large`. The left-hand sub-figures show the full process with both merge phase (shown in blue) and polishing phase (shown in red). Figure 8a demonstrates the power of intensification of the search with the sliding window heuristic. Here, the blue merge plot can seen to have effectively converged as there has been no significant improvement in objective value for over half an hour of wall time. Applying the sliding window heuristic to this "converged" solution results in an almost immediate improvement, followed by subsequent improvements, before this too converges to its local optimum.

Not as emphatic in demonstrating this point are the results presented in Figure 8c. It can be seen in this figure that merge search was still a ways off from converging, although it had started to slow its progress. It could be argued that, were merge search allowed to continue, it would have found the same solution eventually. Although, it can be seen in the plot that applying the polishing heuristic to this unconverged solution did still result in an initial rapid increase in solution quality; and as merge search was beginning to tail off, it probably would have taken longer to reach its goal.

The results from these experiments suggest that, while merge search is adept at finding a good quality global solution, it benefits from the addition of a more exhaustive local method to intensify its search.

Updated state-of-the-art results: When this research was first undertaken, the current published state-of-the-art results were indeed those available on the *minelib* website. However, since then, there have been several developments as published in a recent paper by [**?**] that has collated all reported improvements to the *minelib* dataset.

According to this aggregation of results, recently there are three separate studies that have improved the most on the *minelib* results for the CPIT problem instances used for this research [**?**], [**?**], [**?**]. These improved results are given in Table VI in terms of their LP gap, and are compared with the original *minelib* results and the polished merge results from earlier in this section.

TABLE VI: Recent updates to state-of-the-art results for *minelib* CPIT instances. Given is the percentage gap between the LP upper bound and the best result achieved for each respective study, compared with the original *minelib* results and the best results from this research.

				new results	
Instance	LP UB	*minelib*	polished merge	publication	gap
`newman1`	2.449E+07	4.12%	**1.26%**	[**?**]	**1.26%**
`zuck_small`	8.542E+08	7.67%	0.83%	[**?**]	**0.71%**
`kd`	4.095E+08	3.08%	**0.10%**	[**?**]	0.14%
`zuck_medium`	7.106E+08	13.40%	5.88%	[**?**]	**5.24%**
`marvin`	8.639E+08	5.00%	0.87%	[**?**]	**0.64%**
`zuck_large`	5.739E+07	1.05%	**0.17%**	[**?**]	0.24%

Although the results from this study do not improve on every known bound in the literature, they certainly are competitive with the state-of-the-art and actually do improve the current best-known bound for the `kd` and the `zuck_large` instances.

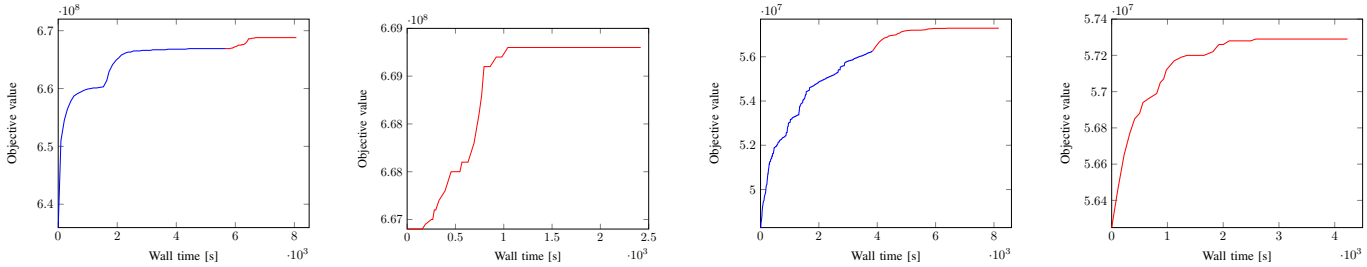
B. The Steiner tree problem in graphs:

Table VII provides the mean objective value and standard deviation for the experiments performed to compare the merge search algorithm to the three base-line algorithms of GRASP, pure local search and pure MIP. The best results for each problem instance are indicated in bold; and, as the optimal solution is known for each instance, if the algorithm managed to find the optimal solution in any of its runs, this is indicated by an asterisk — clearly, if a result has an asterisk and a standard deviation of 0.00, this indicates the optimal solution was found in every run.

The results for the B dataset show that merge search demonstrably outperformed all of the algorithms it was tested against, as do the rest of the datasets. In this first dataset comprised of all the smallest sized instances (around 60 to 290 decision variables, pre-processed), merge search performed the best; managing to find the optimal solution in every run for every problem instance, with many terminating in the first couple of iterations. The next most successful algorithm was the pure MIP algorithm[3], which just edges out pure local search and finally followed by GRASP at the end. The fact that both pure local search and pure MIP outperformed GRASP suggests that the quality of the initial solution is very important to the quality of the resultant solution — especially in these smaller problem instances — so, as pure local search and pure MIP both use the same, deterministic, solution construction as merge search, they are likely to perform better than GRASP, which uses the random solution construction heuristic.

The picture begins to change slightly with the results for the C dataset. Merge search is still the clear winner of the

[3]It should be pointed out here that the pure MIP algorithm failed to produce even a single integer solution for most of the problem instances in the time allotted to it, if it was not given a heuristically constructed solution as a warm start. Therefore the success of the pure MIP algorithm over the GRASP algorithm for this dataset should be attributed to the fact that the MIP search was initially seeded with a higher quality solution than the GRASP algorithm.



(a) Merge with polishing on `zuck_medium`

(b) Detail of polishing phase

(c) Merge with polishing on `zuck_large`

(d) Detail of polishing phase

Fig. 8: Plot of merge search with solution polishing on `zuck_medium` (top) and `zuck_large` (bottom) instances. The left-hand figures show the entire merge search (blue) and polishing (red) phases; the right-hand figures show details of the polishing phases.

TABLE VII: Results on *steinlib* dataset B, C, D and E instances. Given are the known optimal solutions and mean objective values and standard deviations for pure local search, pure MIP, GRASP heuristic and merge search heuristic. Algorithms which managed to find the optimal solution in at least one of its runs are indicated by * next to the objective value.

		pure LS		pure MIP		GRASP		merge search	
Inst.	Opt.	μ	σ	μ	σ	μ	σ	μ	σ
b01	82	**82.00***	0.00	**82.00***	0.00	83.20*	2.40	**82.00***	0.00
b02	83	**83.00***	0.00	**83.00***	0.00	**83.00***	0.00	**83.00***	0.00
b03	138	**138.00***	0.00	**138.00***	0.00	**138.00***	0.00	**138.00***	0.00
b04	59	**59.00***	0.00	**59.00***	0.00	59.80*	1.60	**59.00***	0.00
b05	61	**61.00***	0.00	**61.00***	0.00	**61.00***	0.00	**61.00***	0.00
b06	122	124.00	0.00	**122.00***	0.00	123.80*	1.60	**122.00***	0.00
b07	111	**111.00***	0.00	**111.00***	0.00	**111.00***	0.00	**111.00***	0.00
b08	104	**104.00***	0.00	**104.00***	0.00	**104.00***	0.00	**104.00***	0.00
b09	220	**220.00***	0.00	**220.00***	0.00	**220.00***	0.00	**220.00***	0.00
b10	86	**86.00***	0.00	88.85*	4.29	**86.00***	0.00	**86.00***	0.00
b11	88	**88.00***	0.00	**88.00***	0.00	89.20*	0.98	**88.00***	0.00
b12	174	**174.00***	0.00	**174.00***	0.00	**174.00***	0.00	**174.00***	0.00
b13	165	167.00*	2.45	**165.00***	0.00	168.20*	1.60	**165.00***	0.00
b14	235	236.00	0.00	235.50*	0.74	239.40	1.20	**235.00***	0.00
b15	318	**318.00***	0.00	**318.00***	0.00	322.20	1.47	**318.00***	0.00
b16	127	130.00*	2.45	134.30*	3.80	130.40*	3.38	**127.00***	0.00
b17	131	**131.00***	0.00	131.25*	0.43	131.40*	0.80	**131.00***	0.00
b18	218	**218.00***	0.00	220.05*	1.99	219.40*	1.02	**218.00***	0.00
d01	106	**106.00***	0.00	106.95*	0.22	106.20*	0.40	**106.00***	0.00
d02	220	**220.00***	0.00	237.00	0.00	226.00*	7.35	**220.00***	0.00
d03	1565	1575.20	1.60	1653.00	0.00	1611.80	9.58	**1568.27**	1.14
d04	1935	1941.00	1.26	2008.00	0.00	1976.40	5.57	**1935.36***	0.64
d05	3250	3253.80	0.75	3311.00	0.00	3293.20	7.57	**3252.50**	0.92
d06	67	69.40*	1.20	71.65	2.20	69.20*	1.83	**67.60***	1.20
d07	103	**103.00***	0.00	104.20*	0.98	**103.00***	0.00	**103.00***	0.00
d08	1072	1090.00	3.63	1147.00	0.00	1129.40	11.88	**1077.70**	2.05
d09	1448	1464.20	4.07	1543.00	0.00	1517.40	13.71	**1450.80***	1.60
d10	2110	2121.00	1.79	2161.00	0.00	2185.60	11.89	**2113.60**	1.11
d11	29	29.40*	0.49	31.60*	0.92	31.00	1.10	**29.00***	0.00
d12	42	**42.00***	0.00	**42.00***	0.00	**42.00***	0.00	**42.00***	0.00
d13	500	510.40	2.06	533.00	0.00	531.80	8.01	**504.20**	1.54
d14	667	677.80	0.75	703.00	0.00	709.40	1.36	**671.50**	0.81
d15	1116	1127.20	1.47	1150.00	0.00	1194.80	9.39	**1121.70**	1.55
d16	13	13.40*	0.49	15.25*	0.99	13.20*	0.40	**13.00***	0.00
d17	23	**23.00***	0.00	24.35*	0.91	24.20*	0.98	**23.00***	0.00
d18	223	239.00	1.41	255.00	0.00	274.40	7.63	**234.80**	0.98
d19	310	335.00	0.63	347.00	0.00	393.60	9.44	**326.10**	1.81
d20	537	543.20	0.40	545.00	0.00	646.80	8.98	**541.70**	0.78

		pure LS		pure MIP		GRASP		merge search	
Inst.	Opt.	μ	σ	μ	σ	μ	σ	μ	σ
c01	85	**85.00***	0.00	86.90*	1.41	**85.00***	0.00	**85.00***	0.00
c02	144	**144.00***	0.00	**144.00***	0.00	146.40*	2.94	**144.00***	0.00
c03	754	756.40*	1.62	780.00	0.00	767.40	4.27	**754.50***	1.12
c04	1079	1084.40	1.85	1115.00	0.00	1095.40	4.76	**1081.92**	0.95
c05	1579	1580.40	0.49	1603.00	0.00	1595.40	5.71	**1579.67***	1.11
c06	55	**55.00***	0.00	58.00*	2.26	56.40*	2.80	**55.00***	0.00
c07	102	102.40*	0.49	114.10	2.14	105.40*	1.74	**102.00***	0.00
c08	509	513.40	1.02	528.00	0.00	528.40	3.20	**511.67**	0.62
c09	707	711.00	0.89	728.00	0.00	724.80	2.14	**708.17***	0.90
c10	1093	1095.40	0.49	1127.00	0.00	1116.00	4.86	**1094.75**	1.09
c11	32	33.00*	0.63	34.90	0.44	32.40*	0.49	**32.00***	0.00
c12	46	46.20*	0.40	48.90	0.44	46.80*	1.17	**46.08***	0.28
c13	258	263.00	1.26	276.00	0.00	269.80	3.31	**260.09***	1.08
c14	323	327.40	1.36	337.00	0.00	336.40	2.94	**324.55**	0.99
c15	556	556.80*	0.40	567.00	0.00	582.40	3.83	**556.00***	0.00
c16	11	12.00	0.00	12.00	0.00	12.40	0.80	**11.00***	0.00
c17	18	19.00	0.00	19.60	0.49	19.80	0.75	**18.00***	0.00
c18	113	120.00	0.89	128.00	0.00	128.40	1.36	**117.64**	0.88
c19	146	153.00	0.89	160.00	0.00	164.40	2.94	**151.09**	0.67
c20	267	268.00	0.00	268.00	0.00	298.20	5.04	**267.91***	0.29
e01	111	**111.00***	0.00	124.70	0.90	**111.00***	0.00	**111.00***	0.00
e02	214	226.60	1.85	234.30	6.29	214.40*	0.80	**214.00***	0.00
e03	4013	4066.60	6.34	4238.00	0.00	4194.80	16.68	**4032.00**	3.41
e04	5101	5138.60	5.68	5310.00	0.00	5292.20	14.52	**5108.60**	2.87
e05	8128	8157.40	3.07	8321.00	0.00	8298.00	21.46	**8134.70**	2.65
e06	73	74.00*	2.00	82.85	2.54	75.00*	4.00	**73.00***	0.00
e07	145	148.40*	3.20	165.65	3.53	148.60*	3.88	**145.90***	1.14
e08	2640	2700.00	5.97	2818.00	0.00	2909.20	29.31	**2658.50**	2.77
e09	3604	3660.40	7.99	3795.00	0.00	3938.60	15.45	**3620.50**	3.14
e10	5600	5649.00	6.42	5760.00	0.00	5996.80	9.99	**5613.70**	2.15
e11	34	34.60*	1.20	39.00	0.00	35.40*	1.74	**34.00***	0.00
e12	67	68.40	0.49	71.00	0.00	68.60*	1.85	**67.40***	0.80
e13	1280	1329.60	5.95	1382.00	0.00	1508.60	20.26	**1302.80**	3.06
e14	1732	1775.00	2.00	1828.00	0.00	2059.00	30.13	**1747.10**	2.26
e15	2784	2824.20	1.94	2867.00	0.00	3351.20	104.52	**2800.00**	1.90
e16	15	16.20	0.40	19.00	0.00	16.80*	1.47	**15.00***	0.00
e17	25	25.80*	0.40	28.00	0.00	27.60*	2.42	**25.00***	0.00
e18	564	626.60	2.06	651.00	0.00	821.00	13.13	**600.70**	2.28
e19	758	803.00	2.83	814.00	0.00	1108.00	20.01	**785.20**	1.47
e20	1342	1356.20	0.40	1358.00	0.00	1969.20	25.67	**1353.40**	0.66

four here; although it has not managed to achieve the optimal solution in every single one of its runs and has actually failed to reach the optimal solution in a few of the instances. The big change here is the fact that pure MIP has fallen to last place for almost every single problem instance, even with the warm start seeding; and by the D dataset it is last in every problem instance. This suggests that, by the time the problem instances reach even this moderate size (between 400 and 5,280 decision variables), the problem is too large to be solved by a MIP solver alone, using comparable computing resources. This is evidenced by the fact that the standard deviation is 0.00 for many of the instances, suggesting that it had a hard time finding a better solution than the one it was seeded with.

It is worth mentioning that some of the state-of-the-art results in the *steinlib* library have been produced by a combination of pre-processing and integral LP formulation, so this does not mean that these results suggest the problems are too large for MIP solvers in general, merely that they are too large for the simple formulation used for this research. Better, more complicated MIP formulations for the STPG are available [**?**], [**?**], [**?**]; however, these experiments are not intended to advance the state-of-the-art for STPG solvers, but to demonstrate the application of merge search to the STPG and that there is a non-trivial benefit that can be gained from doing so. If both

pure MIP and merge search used better formulations, and more sophisticated pre-processing techniques, they would indeed be able to achieve better quality solutions; but one could expect a similar gulf in quality between the pure MIP and the hybrid meta-heuristic to emerge — albeit, on much larger problem instances.

The trend of pure local search outperforming the GRASP heuristic continues throughout the rest of the datasets, with the exception of problem instances `c11`, `d06`, `d16` and `e02`. Differences between the mean objective values for these instances are small enough to suggest that these are anomalous and probably the result of the random construction heuristic finding a better initial solution to the deterministic heuristic, or simple luck during the local search.

The fact that pure local search outperformed GRASP in nearly every problem instance suggests that the choice of initial solution is very important to the quality of the solution produced, as otherwise, these two algorithms are nearly identical; so the use of the deterministic construction heuristic is preferable to the random one. Added to this, the fact that merge search equals — or outperforms — pure local search in every problem instance suggests that there is indeed a benefit to creating a hybrid meta-heuristic using a MIP solver that solves a reduced sub-problem induced by merging a population of solutions produced by the local search neighbourhood. Finally, the fact that the merge search algorithm equals — or outperforms — the pure MIP algorithm in every problem instance suggests that these benefits that are gained are not simply from the addition of a MIP solver itself; and therefore, must be the result of the hybridisation and population merging processes.

TABLE VIII: Representative sample of results across *steinlib* dataset instances, investigating the difference between the deterministic and random construction heuristics. Given are the average objective value and standard deviation for solutions produced by both heuristics over 30 runs and the difference between the averages, expressed as a percentage of the larger value.

	deterministic		random		
Instance	μ	σ	μ	σ	difference
b01	82.00	0.00	103.00	13.70	20.39 %
b08	104.00	0.00	141.15	20.57	26.32 %
b16	137.00	0.00	251.35	32.95	45.49 %
b18	226.00	0.00	339.90	45.29	33.51 %
c01	88.00	0.00	289.60	86.30	69.61 %
c08	528.00	0.00	971.40	193.92	45.65 %
c16	12.00	0.00	142.20	51.42	91.56 %
c20	268.00	0.00	725.65	23.83	63.07 %
d01	107.00	0.00	492.70	145.48	78.28 %
d08	1,147.00	0.00	1,929.80	302.90	40.56 %
d16	16.00	0.00	237.60	60.66	93.27 %
d20	545.00	0.00	1,486.45	119.80	63.34 %
e01	125.00	0.00	785.75	280.13	84.09 %
e08	2,818.00	0.00	4,404.75	111.45	36.02 %
e16	19.00	0.00	415.95	152.94	95.43 %
e20	1,358.00	0.00	3,969.15	502.32	65.79 %

TABLE IX: Representative sample of results across *steinlib* dataset instances, investigating the effect that population size has on the number of partitions in the reduced sub-problem. Given are the size of the instance ($|V| + |E|$) and average number of partitions and standard deviation over 30 runs for population sizes 50, 100, 1,500 and 10,000.

		50		100		1,500		10,000	
Instance	size	μ	σ	μ	σ	μ	σ	μ	σ
b01	57	12.40	2.06	15.60	0.80	19.60	0.49	20.00	0.00
b07	93	13.60	1.74	18.60	2.94	29.40	0.49	30.00	0.00
b14	148	43.80	2.79	56.40	3.50	76.20	0.40	77.00	0.00
b18	282	54.60	7.20	82.20	5.98	155.60	2.42	163.60	1.20
c01	421	19.80	2.23	19.20	1.72	28.00	0.63	28.80	0.40
c07	1,259	39.80	3.49	50.20	5.42	70.40	2.15	68.40	1.85
c14	2,695	96.60	7.68	176.20	8.33	912.80	14.76	1,377.00	7.92
c20	5,270	97.60	2.42	188.40	6.15	1,545.80	24.11	2,860.80	27.35
d01	799	20.20	4.96	29.60	0.80	30.60	0.80	30.20	0.40
d07	2,504	25.75	6.61	43.25	7.79	72.20	3.31	78.25	4.82
d14	5,658	93.75	3.63	183.25	4.60	1,426.80	17.59	2,696.75	22.00
d20	11,471	98.00	1.58	201.50	2.96	2,241.00	10.60	5,009.00	27.50
e01	1,972	27.75	7.46	32.75	2.05	37.00	3.41	35.75	2.49
e07	6,277	56.50	7.09	86.50	3.84	231.00	25.79	245.75	15.07
e14	14,617	96.00	2.55	191.25	7.50	2,281.80	22.48	5,835.25	45.78
e20	31,593	97.75	6.02	190.00	5.10	2,951.40	18.10	9,532.50	19.93

Additional experiments: Table VIII illustrates the difference between using the deterministic solution construction heuristic and the random solution construction heuristic. It is very clear to see that the quality of the deterministic heuristic is much better than the average random construction heuristic, especially for the larger datasets, where the difference can be up to over 95%. The full set of results are available in Table **??** of the appendix.

The results in Table IX show the effect that changing the population size has on the number of partitions in the resultant reduced sub-problem, before random splitting is applied. It can be seen from the data that as the size of the population increases, the number of partitions also increases. This is to be expected, as two variables are included in the same partition only if they agree on values across the entire population of solutions; therefore, if more solutions are considered, the less likely variables are to agree across them all and the more partitions in the reduced sub-problem. Again, these are just a representative set of the problem instances, the full results are given in Table **??** of the appendix.

Figure 9 shows some of the plots that were used in the sensitivity analysis done to determine an appropriate population size for these experiments. The number of partitions in the reduced sub-problem was plotted against population size from 0 to 5,000 and a population size of 1,500 was decided upon for two main reasons. The first reason is that it can be seen in the plots that although the number of partitions is proportional to the population size, it is not directly proportional, and after a certain inflection point there are significant diminishing returns in the number of partitions for population size. It can be seen from the plots that the position of this inflection point moves further to the right with the size of the problem and it was decided that 1,500 was a decent value that could be used across the whole dataset. It is slightly after the inflection point for most of the C dataset instances and slightly before the inflection point for most of the D dataset instances. The B dataset instances are small enough that it doesnt matter that 1,500 is probably too many and, although it falls very short of

the E dataset instances inflection point, larger population sizes produce too many partitions for the MIP solver to handle.

The second reason for choosing this population size is that in order for the MIP solver to be effective, the number of decision variables must be kept at a reasonable amount. It was decided that the number of partitions in the reduced sub-problem, after random splitting, should a maximum of around 2,000, on average. The number of random splits was set at 500, therefore the number of partitions before splitting should be around a maximum of 1,500 — which occurs at a population size of around 1,500 for many of the problem instances. For larger problem instances which need a larger reduced sub-problem, an extra parameter was added M_{factor}, which ensures that the reduced sub-problem is some fraction of the total instance size.

VI. Conclusion and future work

Constrained optimisation problems involve the search for an optimal object in (often *very* large) search spaces — in the case of combinatorial optimisation, these spaces are finite and discrete. As well as this, problems that are abstractions of real-world processes can be extremely complex; having many additional side-constraints, separate to the main problem constraints themselves, which must be considered when modelling the problem mathematically.

[**?**] showed that these ($\mathcal{NP}$-Complete) problems are all theoretically reducible to one another; but in practice, they often require specific domain knowledge that means custom algorithms must be developed for each problem, individually. The discovery of a "one size fits all" solution for this whole class of problems would be a vital achievement in the field of operations research, and the purpose of this thesis was to attempt to provide a material step in the direction towards this goal.

Merge search, the algorithm proposed here, is a generalised hybrid meta-heuristic framework for solving large-scale and complex constrained optimisation problems — with, or without specific domain knowledge. The operation of this framework can be broken down into several phases:

- *Initial solution construction:* a feasible solution to the problem is constructed, either heuristically or by solving a generic mixed integer program (MIP) model of the problem that penalises non-feasible solutions;
- *Population generation:* a population of neighbouring solutions is generated, using some local search heuristic or by randomly perturbing an initial solution;
- *Partitioning variables:* the population of solutions is used to determine how the set of decision variables are partitioned. Additional splits to these partitions can be made to increase the granularity of the solutions produced; and,
- *Solving reduced sub-problem:* new local optima can be found by solving a reduced version of the problem which uses each partition as a meta-variable, either with a MIP solver or by some other method. By grouping the decision variables, this sub-problem requires much less computational effort to solve — the trade-off being that the solution produced becomes an approximation of the optimal one.

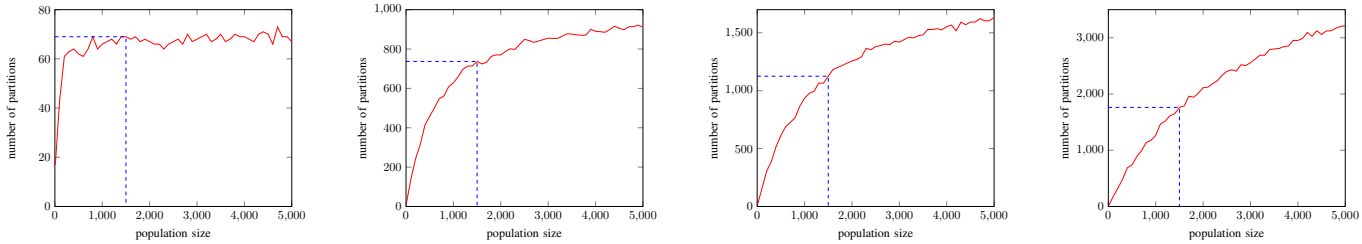
A theoretical description of merge search was given along with a generic version which, although not as efficient as methods which utilise *some* domain-specific knowledge, would serve to operate as a black-box solver for any combinatorial optimisation problem, provided a MIP model exists for it. It was also shown that merge search can be considered a generalisation of many optimisation techniques such as crossover and mutation operators in genetic algorithms, or search techniques like large neighbourhood search.

Practically, two problems were used as testing-grounds to demonstrate the effectiveness of merge search and explore some of its features and facets. The first was a complex, real-world problem from the field of open-pit mining called the constrained pit-limit (CPIT) problem and the second was a well-known problem from Karp's original 21 $\mathcal{NP}$-Complete problems, the Steiner tree problem in graphs (STPG). These two problems were chosen because they are very dissimilar and require very different approaches to solve them, which allowed the versatility and flexibility of this hybrid meta-heuristic framework to be showcased.

As it is a much simpler problem, the experiments on the STPG were used to investigate some of the main properties and design choices that need to be made when constructing merge search algorithms, specifically solution construction methods, population size, method of variable partitioning and methods of solving the reduced sub-problem. In contrast, the experiments on the CPIT problem highlighted the effectiveness of merge search in solving large scale, highly constrained, problems through its automatic decomposition aspects. They also investigated the effect that additional arbitrary splitting of the merge partitions has on the quality of the solutions produced.

While the datasets used for the STPG experiments already had been solved to optimality, and therefore were only being used to demonstrate the versatility of merge search and investigate its various aspects; the experiments on the CPIT problem produced solutions which improved on the upper-bounds of all six of the problem instances, as published on the *minelib* website, and two of the most recent published upper-bounds as reported in [**?**].

Although demonstrating that merge search can be applied to two very different problems shows that the algorithm is very versatile, making the claim that it is a general framework that can be applied to *any* combinatorial optimisation problem would require significantly more evidence than only two problems. Therefore, the most obvious future research that can be performed is in applying the technique to many more (and varying types of) problems. Aside from this, some further work could be done into developing a truly generic algorithm which operates entirely on the decision variables, and is completely problem agnostic; the interaction of different heuristics and other techniques, within the framework; the effectiveness of merge search with other exact solvers (such as constraint programming); the limits at which merge search is



(a) Instance `b10` (b) Instance `c10` (c) Instance `d10` (d) Instance `e10`

Fig. 9: Effect of population size on number of partitions in the reduced sub-problem. Dotted line shows population size of 1,500, which was used for these experiments.

most effective; and, applying merge search to different classes (possibly non-discrete) of problems.

APPENDIX A
PROOF OF THE FIRST ZONKLAR EQUATION

Appendix one text goes here.

APPENDIX B

Appendix two text goes here.

ACKNOWLEDGMENT

The authors would like to thank...

REFERENCES

[1] C. H. Papadimitrou and K. Steiglitz, "Combinatorial optimization: algorithms and complexity," 1982.
[2] K. Deb, *Optimization for engineering design: Algorithms and examples*. PHI Learning Pvt. Ltd., 2012.
[3] C. Blum, J. Puchinger, G. R. Raidl, and A. Roli, "Hybrid metaheuristics in combinatorial optimization: A survey," *Applied Soft Computing*, vol. 11, no. 6, pp. 4135–4151, 2011.
[4] L. A. Wolsey and G. L. Nemhauser, *Integer and combinatorial optimization*. John Wiley & Sons, 1999, vol. 55.
[5] G. B. Dantzig and P. Wolfe, "Decomposition principle for linear programs," *Operations research*, vol. 8, no. 1, pp. 101–111, 1960.
[6] M. N. Omidvar, X. Li, Y. Mei, and X. Yao, "Cooperative co-evolution with differential grouping for large scale optimization," *IEEE Transactions on evolutionary computation*, vol. 18, no. 3, pp. 378–393, 2013.
[7] D. Bertsimas and J. N. Tsitsiklis, *Introduction to linear optimization*. Athena Scientific Belmont, MA, 1997, vol. 6.
[8] R. Chicoisne, D. Espinoza, M. Goycoolea, E. Moreno, and E. Rubio, "A new algorithm for the open-pit mine production scheduling problem," *Operations Research*, vol. 60, no. 3, pp. 517–528, 2012.
[9] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
[10] A. M. Newman, E. Rubio, R. Caro, A. Weintraub, and K. Eurek, "A review of operations research in mine planning," *Interfaces*, vol. 40, no. 3, pp. 222–245, May 2010.

PLACE PHOTO HERE

Michael Shell Biography text here.

John Doe Biography text here.

Jane Doe Biography text here.