Learning Bayesian Networks for Large Domains

Walter Perez Urcia

A dissertation proposal submitted
to
Instituto de Matemática e Estatística
of
Universidade de São Paulo
in partial fulfillment
of the requirements for
Master of Science
Degree in
Computer Science

Program: Computer Science
Advisor: Prof. Dr. Denis D. Mauá

This work received financial support from CAPES

São Paulo, December 2016

Learning Bayesian Networks for Large Domains

Esta é a versão original da dissertação elaborada pelo
candidato Walter Perez Urcia, tal como
submetida à Comissão Julgadora.

Acknowledgements

We thank Mauro Scanagatta for providing us with some of the data sets used. This work was developed with the help of HPC resources available by Superintendencia de Informação of Universidade de São Paulo.

Resumo

Perez Urcia, Walter. **Aprendizado de Redes Bayesianas para domínios grandes**. 2017, 38 f. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2017.

Redes Bayesianas são modelos gráficos amplamente utilizados para automatizar o raciocínio probabilístico em domínios complexos. Uma rede Bayesiana é um grafo direcionado acíclico no qual os nós representam variáveis aleatórias e os arcos representam relações de dependência entre variáveis. Especificar manualmente uma rede Bayesiana sobre um domínio grande e complexo é uma tarefa altamente custosa e propensa a erros. Isto justifica o desenvolvimento de métodos para aprender estruturas de redes Bayesianas a partir de observações.

Uma abordagem bem sucedida para aprendizado de redes Bayesianas é especificar uma função de pontuação (score), que associa cada estrutura a um número que representa a adequação do modelo aos dados e ao conhecimento prévio. O aprendizado então consiste em selecionar uma estrutura com alta pontuação. A aprendizagem estrutural baseada em pontuação é uma tarefa computacionalmente custosa (NP-difícil), o que cria a necessidade de desenvolvimento de técnicas aproximadas. Embora existam muitas técnicas com alguma garantia de qualidade (convergência, consistência ou estimativa de erro), elas possuem um custo computacional alto e não são aplicáveis em domínios muito grandes (centenas ou milhares de variáveis).

Uma técnica simples e eficaz para a aprendizagem aproximada de estruturas de redes Bayesianas consiste em realizar uma busca local no espaço de ordenações topológicas de variáveis utilizando um espaço restrito de conjuntos de pais. Embora essa abordagem não possua garantias de desempenho, ela é computacionalmente eficiente, e empiricamente superior a outros métodos, especialmente quando o número de variáveis é grande. Geralmente, a busca local é inicializada com uma ordenação das variáveis gerada uniformemente no espaço de ordenações. Isso pode levar a busca a obter soluções de baixa qualidade e a requerer um número alto de iterações, o que prejudica o desempenho do método.

Esse trabalho tem como objetivo estudar e aprimorar as técnicas de aprendizagem de redes Bayesianas em domínios muito grandes. Em particular, pretende-se melhorar a qualidade das soluções encontradas pelo algoritmo de busca por geração de ordenações topológicas, empregando técnicas do estado-da-arte na geração de conjuntos de pais e desenvolvendo heurísticas informadas para geração de ordenações topológicas. A qualidade das soluções encontradas foi avaliada pela pontuação. Os resultados mostram que as novas heurísticas de inicialização melhoram as redes obtidas com uma diferencia significativa.

Palavras-chave: Redes Bayesianas, Aprendizagem de máquina, Modelos probabilísticos, Busca local.

Abstract

Perez Urcia, Walter **Learning Bayesian Networks for Large Domains**. 2017, 80 f. Dissertatation (Master Degree) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2017.

Bayesian networks are widely used graphical models for reasoning under uncertainty on complex domains. A Bayesian network is a directed acyclic graph where nodes represent random variables and the arcs represent (in)dependence relationships. Manually specifying a Bayesian network over a large and complex domain is a time-consuming and error-prone task. This justifies the development of methods for learning Bayesian network structures from data.

A successful approach to Bayesian network structure learning is to use a score function which assigns a value for each structure based on how well the structure represents the data. This way the problem of learning a Bayesian network becomes a combinatorial optimization of finding structures. Score-based structure learning is a computationally demanding task (in fact, NP-Hard), which justifies the development of approximate methods. Even though there are some methods which provide quality guarantees (convergence, consistence or error estimative), they scale poorly to large domains (hundreds and thousands of variables).

An effective approach for learning Bayesian network structures is to perform a local search on the space of topological orderings using a restricted space of parent sets. While this approach has no performance guarantee, it is computationally efficient and performs empirically better than other approaches, especially on large domains. Typically, the search is initialized with a randomly generated ordering. This can lead to poor local optima, slow convergence and ultimately degrade the performance of the method as the number of variables increases.

This work aims at studying and improving order-based local search methods for score-based Bayesian network structure learning on large domains. Specifically, we aim at improving solutions obtained by order-based local searches using state-of-the-art parent set selection methods, and at developing new informed heuristics that allow for learning better large Bayesian networks. The new heuristics were evaluated on the scores obtained from real-world data sets. Results show that the new initialization heuristics improve the obtained Bayesian networks significantly.

Keywords: Bayesian networks, Machine Learning, Probabilistic Model, Local Search.

Contents

List of Symbols **ix**

List of Figures **xi**

List of Tables **xiii**

1 Introduction **1**

1.1 Objectives . . . 2
1.2 Contributions . . . 2
1.3 Organization . . . 3

2 Background **5**

2.1 Learning Bayesian networks . . . 5
2.2 Parent Set Selection . . . 7
2.2.1 Safe Pruning Rules . . . 7
2.2.2 Generic algorithm for parent set selection . . . 7
2.2.3 Sequential Selection . . . 8
2.2.4 Greedy Selection . . . 9
2.2.5 Independence Selection . . . 9
2.2.6 AD-Tree . . . 10
2.3 Order-based structure optimization . . . 10
2.3.1 Greedy Search . . . 11
2.3.2 Simulated Annealing . . . 11
2.3.3 Tabu Search . . . 12
2.3.4 Beam Search . . . 12
2.3.5 Acyclic selection order-based search . . . 13
2.3.6 Swap Search . . . 14

3 Contributions **17**

3.1 Generating informed initial solutions . . . 17
3.2 Best Parent Set Graph . . . 17
3.3 DFS-based approach . . . 17
3.4 FAS-based approach . . . 19
3.5 Best-first approach . . . 20

4 Results and Discussion **23**
4.1 Experimental Setup . . . 23
4.2 Parent set selection approaches . . . 23
4.3 Informed initialization heuristics . . . 25
4.4 Structure optimization methods . . . 25
4.4.1 Greedy Search . . . 25
4.4.2 Simulated Annealing . . . 27
4.4.3 Tabu Search . . . 27
4.4.4 Beam Search . . . 28
4.4.5 Acyclic Selection Order-based Search . . . 29
4.4.6 Swap Search . . . 30
4.5 Discussion . . . 31

5 Conclussions and Future Work **33**
5.1 Conclusions . . . 33
5.2 Future work . . . 33

A Configuration Parameters For Each Local Search Method **35**
A.1 Greedy search . . . 35
A.2 Simulated annealing . . . 35
A.3 Tabu search . . . 35
A.4 Beam search . . . 35
A.5 ASOBS . . . 35
A.6 Swap search . . . 35

B Empirical Results: Greedy Search **37**

C Empirical Results: Simulated Annealing **41**

D Empirical Results: Tabu Search **45**

E Empirical Results: Beam Search **49**

F Empirical Results: ASOBS **53**

G Empirical Results: Swap Search **57**

Bibliography **61**

List of Symbols

$\mathbf{C}_i^L$ Candidate parent sets for variable X_i consistent with order L. Pages 11, 17

$\mathbf{C}_i$ Candidate parent sets for variable X_i. Pages 7, 10, 11, 20, 21

r_i Number of states variable X_i can assume. Page 5

N Number of instances on the domain. Pages 5–7, 10

n Number of random variables on the domain. Pages 5–7

G^* Optimal Bayesian network. Pages 6, 7

Pa_i^G Parent set for variable X_i in graph G. Page 5

X_i Random variable on the domain. Pages 5, 7, 11, 13, 14

$sc(G)$ Score function value of graph G. Page 5

$sc_i(Pa_i^G)$ Score function value of variable X_i given its parents on graph G. Page 5, 6

$<$ Variable ordering. Pages 6, 7

List of Figures

2.1 Bayesian network for the domain A, B, C, D 6
2.2 Search space sizes (log-scale) for $n = 1, \dots, 25$ 6
2.3 Parental graph of a variable X with respect to domain X, Y, Z, W. 8
2.4 AD-Tree for a domain X, Y, Z. 10
2.5 DAG G acyclicly consistent with order $[A, B, C, D]$ 14

3.1 An example of a best parent set graph. 18

4.1 Nemenyi post-hoc analysis average initial score 25
4.2 Nemenyi post-hoc analysis greedy search + sequential selection 26
4.3 Nemenyi post-hoc analysis greedy search + greedy selection 26
4.4 Nemenyi post-hoc analysis greedy search + independence selection 26
4.5 Nemenyi post-hoc analysis simulated search + sequential selection 27
4.6 Nemenyi post-hoc analysis simulated search + greedy selection 27
4.7 Nemenyi post-hoc analysis simulated search + independence selection 27
4.8 Nemenyi post-hoc analysis tabu search + sequential selection 28
4.9 Nemenyi post-hoc analysis tabu search + greedy selection 28
4.10 Nemenyi post-hoc analysis tabu search + independence selection 28
4.11 Nemenyi post-hoc analysis beam search + sequential selection 28
4.12 Nemenyi post-hoc analysis beam search + greedy selection 29
4.13 Nemenyi post-hoc analysis independence search + independence selection 29
4.14 Nemenyi post-hoc analysis asobs search + sequential selection 29
4.15 Nemenyi post-hoc analysis asobs search + greedy selection 30
4.16 Nemenyi post-hoc analysis asobs search + independence selection 30
4.17 Nemenyi post-hoc analysis swap search + greedy selection 30
4.18 Nemenyi post-hoc analysis swap search + independence selection 31
4.19 Nemenyi post-hoc analysis with best setups 31

List of Tables

2.1 Effect of pruning rules . . . 8

3.1 Possible execution of DFS-based generation . . . 19

4.1 Data sets characteristics . . . 24
4.2 Data sets scores . . . 24
4.3 Best setup for each algorithm . . . 31

B.1 Results using sequential selection . . . 37
B.2 Results using greedy selection . . . 38
B.3 Results using independence selection . . . 39

C.1 Results using sequential selection . . . 41
C.2 Results using greedy selection . . . 42
C.3 Results using independence selection . . . 43

D.1 Results using sequential selection . . . 45
D.2 Results using greedy selection . . . 46
D.3 Results using independence selection . . . 47

E.1 Results using sequential selection . . . 49
E.2 Results using greedy selection . . . 50
E.3 Results using independence selection . . . 51

F.1 Results using sequential selection . . . 53
F.2 Results using greedy selection . . . 54
F.3 Results using independence selection . . . 55

G.1 Results using sequential selection . . . 57
G.2 Results using greedy selection . . . 58
G.3 Results using independence selection . . . 59

Chapter 1

Introduction

Bayesian Networks are space-efficient representations of multivariate probability distributions [Pea88]. They are defined by two components: an acyclic directed graph (DAG), known as the *structure* which encodes the independencies among the variables, and a collection of conditional probability distributions of each variable given its parents.

Bayesian networks formalize reasoning under uncertainty on complex domains. A Bayesian network has two main advantages: i) it is a complete model in that it specifies a unique joint probability distribution over the network variables, and ii) it allows for a compact and intuitive representation of the knowledge about the domain [Dar09].

Manually specifying a Bayesian network is a difficult and time-consuming task. That is why practitioners often resort to *learning* the model from data.

Bayesian network learning is usually decomposed into two steps: first a DAG is obtained (structure learning), and then the numerical parameters (i.e., the conditional probabilities) are estimated for a fixed DAG (parameter learning). When data is complete (i.e., there are no missing or corrupted values), parameter learning boils down to counting and can be performed efficiently. This reduces the problem of learning a Bayesian network into the problem of learning a structure which satisfies some optimality criterion.

One approach to structure learning consists in performing multiple conditional independence tests while enforcing acyclicity [SM95, CGK+02]. Although this approach, known as *constraint-based*, is consistent in that it recovers the "true" DAG (if one exists) given infinite data and computational resources, testing for independency can be computationally demanding and highly sensitive to the significance level used; this last problem is particularly important in small-sample or high dimensional data sets [HMC06].

Another approach, known as *score-based*, consists in associating every DAG with a polynomial-time computable score value. Typical score functions reward DAGs with high probability of generating the data set (i.e., the data likelihood) while penalizing the complexity of the model (i.e., the number of parameters). Some examples are the Bayesian Information Criterion (BIC) [Sch78], the Akaike Information Criterion (AIC) [Aka74], Minimum Description Length (MDL) [CT91] and the Bayesian Dirichlet score (BD) [HGC95].

There are also hybrid approaches that employ statistical testing to generate good candidate structures and then uses this as a starting point for a score-based approach [TBA06]. The hybrid approach alleviates the false positive problem at an increased computational cost.

Chickering showed that score-based learning is NP-hard for a wide class of score functions [Chi96]. This result was later strengthened to show NP-hardness for large-sample domains [CHM04]. Note that the latter includes the case of hybrid learning methods. Those results support the development of approximate methods that can scale for large domains.

An effective approach for score-based Bayesian network structure learning is to resort to local search methods that find an approximate solution; for example, by searching over the space of DAG structures or score-equivalent graphs [CH92, LB94, FNP99, Chi02]. Based on a common observation that score-based structure learning is tractable when an ordering over the variables and the number of parents are fixed [Bun91], Tessyer and Koller [TK05] developed a local search method, known as *order-based local search* (OBS), that remains as one of the most competitive in terms of time efficiency and accuracy. Recently, Scanagatta et. al. [SdCCZ15] extended OBS in order to improve its accuracy with a small increase in the complexity. This later approach, known as *acyclic selection OBS* (ASOBS), was shown empirically be the state-of-the-art in structure learning.

As with most local search approaches, one can infer that the choice of a good initialization in (AS)OBS is crucial to the quality of the solution found. However, the search is usually initialized with an ordering sampled uniformly at random. This can lead to poor local optima, slow convergence, and hurt the performance of the method as the number of variables increases. These problems can be alleviated by employing more sophisticate techniques such as *tabu search* [Glo89], *simulated annealing* [GKR94] or *data perturbation* [ENFS02], but this usually adds a significant computational overhead. A less demanding solution is to use a better initialization heuristic, that is, a function that returns an initial ordering in a high-scoring region with a low complexity algorithm.

1.1 Objectives

This work aims at providing an extensive comparison of the state-of-the-art order-based learners for *large* Bayesian networks, that is, Bayesian networks with hundreds or thousands of variables. Additionally, this work aims at improving the performance of the solutions obtained (measured by its score) by designing novel informed initialization heuristics. Specifically, we target the performance of the solutions obtained by order-based *local search* approaches using state-of-the-art parent set selection methods and new informed heuristics for generating initial solutions.

1.2 Contributions

The main contributions developed in this work are:

- Three novel initialization heuristics for order-based local search methods

A preliminary version of this work containing the first two heuristics was published at the Proceedings of 3rd Symposium on Knowledge Discovery, Mining and Learning (KDMiLe 2015) [PM15]. That version aimed at providing an statistical analysis of the impact of initialization heuristics and parent set selection methods on the performance of order-based greedy search method. It was presented at the venue by the author at Laboratório Nacional de Computação Científica (LNNC), Petrópolis, RJ.

- A C++ implementation of state-of-the-art order-based structure learning approximated algorithms

The second contribution was developed because the preliminary version showed a necessity for a scalable implementation for large domains which also allows comparison among state-of-the-art structure learning approximated algorithms on a same environment. This new package was used to make an extended version of [PM15]'s work and was accepted for publication on the Journal of Information and Data Management (JIDM).

- An empirical statistical analysis of state-of-the-art order-based approximated algorithms

Finally, the third heuristic was developed for the new state-of-the-art acyclic selection OBS and published on the XIII Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2016) [PM16]. As a by product, we also provide an extensive empirical evaluation of ASOBS (more extensive than the one performed by the authors in [SdCCZ15]).

The new heuristics were evaluated according to the scores obtained from real-world data sets.

1.3 Organization

The rest of this monograph is structured as follows. We begin in Chapter 2 explaining the basic concepts of score-based Bayesian network structure learning. Then, in Chapter 3, the new heuristics for generating initial solutions are described. Chapter 4 contains experimental results and statistical analysis. Finally, in Chapter 5 we state the conclusions and future works.

Chapter 2

Background

In this chapter we describe local-search approaches for score-based Bayesian network structure learning. We begin in Section 2.1 with a review of score-based structure learning. In Section 2.2, we discuss the main techniques for selecting candidate parent sets, a sub-step of structure learning. Then, in Section 2.3, we describe the state-of-the-art order-based approximated algorithms for structure learning.

2.1 Learning Bayesian networks

A Bayesian network is defined by a DAG $G = (V, E)$, where $V = \{X_1, X_2, \ldots, X_n\}$ is the set of (categorical) random variables, and a collection of conditional probability distributions $P(X_i \mid Pa_i^G)$, $i = 1, \ldots, n$, where Pa_i^G are the parents of X_i in G and $P(X_i|\emptyset) = P(X_i)$. The Bayesian network is assumed to induce a joint probability distribution over all the variables through the equation

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pa_i^G).$$

The number of numerical parameters required to specify a Bayesian network with DAG G is $size(G) = \sum_{i=1}^{n}(r_i - 1)\prod_{X_j \in Pa_i^G} r_j$, where r_i denotes the number of states variable X_i can assume. For example, the network in Figure 2.1 has $size(G) = 9$ parameters considering that each variable can assume two possible values.

A *score function* $sc(G)$ assigns a real-value to any DAG G indicating its goodness in representing a given data set.[1] We assume that the score function is *consistent*, that is, if the data set contains N instances generated by a Bayesian network with graph G^* then when $N \to \infty$ it follows that $sc(G^*) > sc(G)$ for any graph G whenever $size(G^*) < size(G)$.

We also require that the score function *decompose*, meaning that it can be written as a sum of local $sc(G) = \sum_{i=1}^{n} sc_i(Pa_i^G)$. Most score functions used in structure learning satisfy these properties, such as the BIC, MDL, and BD [CHM04]. For example, the BIC score function, is given by

$$\mathrm{BIC}(G) = LL(G) - \frac{\ln N}{2} size(G) = \sum_{i=1}^{n} \underbrace{\overbrace{\sum_k \sum_j N_{ijk} \ln \frac{N_{ijk}}{N_{ij}}}^{\text{local data loglikelihood}} - \frac{\ln N}{2} \overbrace{(r_i - 1) \prod_{X_j \in Pa_i^G} r_j}^{\text{local penalization}}}_{\mathrm{BIC}_i(Pa_i^G)}$$

where $LL(G)$ is the data loglikelihood, N_{ijk} is the number of instances where attribute X_i takes its kth value and its parents take the jth configuration (for some arbitrary fixed ordering of the configurations of the parents' values), and similarly for N_{ij}.

[1] The dependence of the scoring function on the data set is usually left implicitly, as for most of this explanation we can assume a fixed data set. We assume in this work that the dataset contains no missing values.

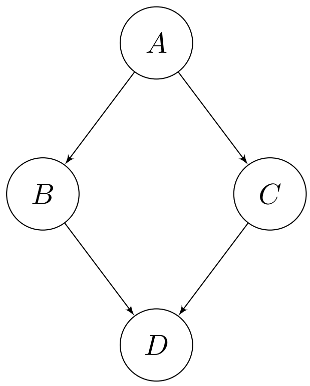


Figure 2.1: *Bayesian network for the domain A, B, C, D*

The Bayesian network structure learning problem is to find G^* that satisfies

$$sc(G^*) = \max_{G \text{ is a DAG}} \sum_{i=1}^{n} sc_i(Pa_i^G). \tag{2.1}$$

Robinson [Rob73] proved that the number of DAGs with n nodes is obtained by the recurrence relation $a_n = \sum_{k=1}^{n}(-1)^{k-1}\binom{n}{k}2^{k(n-k)}a_{n-k}$, which grows super-exponentially.

A DAG always admits a topological ordering of its nodes, and from a topological ordering it is always possible to construct a DAG by selecting as parents of variable a subset of smaller variables. Thus, an alternative approach to structure learning is to search over the space of DAGs which are consistent with a topological ordering, that is, to find DAG G^* such that

$$sc(G^*) = \max_{<} \sum_{i=1}^{n} \max_{\mathbf{Y} \in \{X_j < X_i\}} sc_i(\mathbf{Y}). \tag{2.2}$$

The size of the search space of this formulation of the problem is $t_n = n!2^n$, which is significantly smaller than the search space of DAGs. Figure 2.2 shows the sizes of both spaces for domains of dimension n up to 25. Note that in Equation (2.2) the inner maximizations (parent set selection) are performed independently.

Under the BIC and MDL score functions, the best parent set of any variable (for any fixed ordering) has at most $\log N$ variables (where N is the size of the data set) [dCJ11]. Thus, one obtains the best parent set of a variable in time $O(n^{\log N})$ by an enumerative search. This naive approach however is impractical even for moderately large domains [SdCCZ15]. Koivisto [Koi06] showed that finding the best parent set of a variable is LOGSNP-hard even to approximate for the

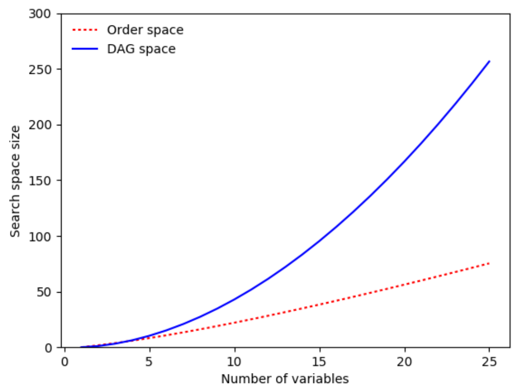


Figure 2.2: *Search space sizes (log-scale) for $n = 1, \ldots, 25$*

MDL, BIC and AIC score functions. The class LOGSNP contains problems that are believed to exhibit runtime $O(s^{\log s})$, where s is the size of the input, yet are thought not be NP-hard (note that $(n^{1+\log N})$ is sub-exponential).

Order-based structure learning is usually performed in two steps: i) first, an exact or approximate technique is used to obtain a list of candidate parent sets $\mathbf{C}_i$ for each variable X_i (i.e., $\mathbf{C}_i$ is a set of subsets of $V \setminus \{X_i\}$); ii) then, these candidate parent sets are used to search for [YM13, BC15, SdCCZ15]

$$\overline{G} \in \arg\max_{<} \sum_{i=1}^{n} \max_{\mathbf{Y} \in \mathbf{C}_i} sc_i(\mathbf{Y}), \tag{2.3}$$

The quality of the learned structure depends on the quality of both steps.

2.2 Parent Set Selection

The theoretical results discussed in the previous section justify the use of approximate techniques for selecting candidate parent sets. As a by-product of this step, we obtain an approximation of the graph H^* which has the locally-optimum parent set for every variable. This graph is the base for the initialization heuristics in Chapter 3.

Obtaining a high-quality list of candidate parent sets $\mathbf{C}_i$ is an important step in effective score-based structure learning problem, but if complete, such a list has 2^{n-1} parent sets for each variable. This justifies the use of pruning rules that exclude some parent sets. Pruning rules that exclude parent sets which are not part of an optimal DAG G^* are known as *safe* pruning rules. *Unsafe* pruning rules such as *sequential*, *greedy* or *independence* selection, rely on some heuristics to select parent sets. In this work, we focus on pruning rules that work on BIC score function.

2.2.1 Safe Pruning Rules

One of the most used pruning rules is to discard any parent set of size greater than k. The following theorem shows a particular application of this rule for the BIC score:

Theorem 1. *[dCZJ09] When using BIC score function, there is no optimal parent set with size greater than* $\log \frac{2N}{\log N}$ *for any variable.*

Thus, discarding any parent set of size greater than $k = \log \frac{2N}{\log N}$ is safe. Another commonly used pruning rule is to prune sets when at least one subset have smaller score. This is represented for the case of BIC score by the following theorem:

Theorem 2. *[dCJ11] Let U and S be two candidate parent sets for variable X such that $U \subset S$, and BIC$(X, U) \geq$ BIC(X, S). Then S or a super set of S is not the optimal parent set of X for any candidate set.*

In order to prune parent sets efficiently, the previous rule can be applied in two different ways: *end-of-calculating* and *while-calculating*. The former consists in checking, after all calculations were performed, for a set S, if at least one subset $U \subset S$ has greater score. The latter consists in checking, during the calculations, whether to expand set S on next iterations when $S = U \cup \{Y\}$ (for some $Y \neq X$).

The application of both safe pruning rules reduces significantly the number of scores $n2^{n-1}$ as shown in Table 2.1. For large domains, the number of candidate parent sets is still too high which justifies the use of unsafe pruning rules.

2.2.2 Generic algorithm for parent set selection

The following unsafe pruning techniques for parent set selection can all be seen as heuristic search approaches in the parental graph of a variable X_i, which is the graph where each node is

Dataset	n	k	Unpruned	Theorem 1	Theorems 1 & 2
Census	15	8	245K	45K (18.33%)	3.5K (1.44%)
Voting	17	4	1.1M	31K (2.78%)	939 (0.08%)
Letter	17	8	1.1M	218K (19.64%)	115K (10.38%)
Hepatitis	17	3	1.1M	9.5K (0.85%)	174 (0.02%)
Image	20	6	10M	542K (5.18%)	6.2K (0.06%)
Heart	23	3	96M	35K (0.04%)	327 (0.00034%)
Mushroom	23	7	96M	3.9M (4.07%)	13K (0.014%)
Parkinsons	23	4	96M	168K (0.17%)	2.4K (0.003%)
Autos	25	4	419M	265K (0.06%)	3K (0.0007%)
Flag	28	4	3.7B	491K (0.01%)	776 (0.00002%)

Table 2.1: *Effect of pruning rules*

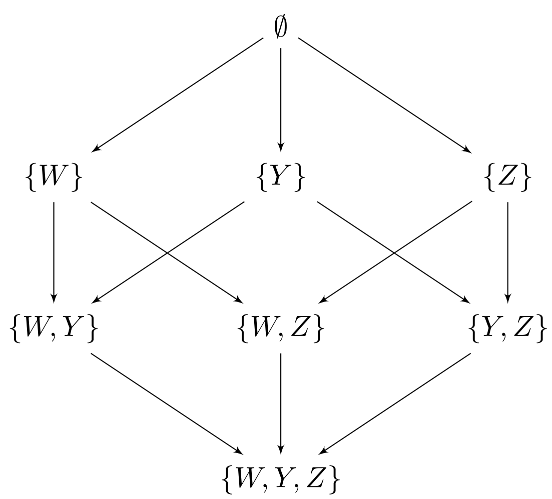


Figure 2.3: *Parental graph of a variable X with respect to domain X, Y, Z, W.*

a subset of $V \setminus \{X_i\}$, and an arc connects a subset/node $\mathbf{Y}$ to a subset/node $\mathbf{Z}$ if $\mathbf{Z} = \mathbf{Y} \cup \{X_j\}$ for some variable $X_j \neq X_i$ [YM13]. For example, Figure 2.3 depicts the parental graph for variable X in a small domain containing variables X, Y, Z, W. Each node $\mathbf{X}$ in the parental graph of X_i is associated with a score $sc_i(\mathbf{X})$. Note that the parental graph of an n-dimensional domain has 2^{n-1} nodes; therefore, exhaustive search is impracticable for large domains.

A general search procedure on the parental graph is showed on Algorithm 1. The functions h_i and g_i denote the heuristic and actual score of a candidate parent set; *open* is a list of parent sets to be explored sorted by the values $h_i(\mathbf{Y})$, and *closed* is a list of parent sets already explored (those for which $g_i(\mathbf{Y})$ has been computed). The techniques differ in terms of the initialization of *open* and *closed* (line 1), and in the definitions of h_i (line 6) and g_i (line 10).

2.2.3 Sequential Selection

Sequential Selection is adopted by the state-of-the-art exact learning algorithms [YM13, BC15]. This technique consists in performing a breadth-first search in the parental graph up to a given depth d. The list *closed* is initialized empty, and *open* is initialized containing the empty set (hence not empty). The function g_i is set as the local score function sc_i, and $h_i(\mathbf{Y}) = -|\mathbf{Y}|$. Pruning rules discussed in subsection 2.2.1 can be applied to detect suboptimal paths during the search [dCJ11]. The worst-case running time of sequential selection is $O(n^d)$ and is optimal (i.e. safe) given a sufficiently large d; however, when working with high-dimensional data sets ($n > 100$), it is necessary to set d to low values (e.g. $d = 2$), which can severely limit the quality of the produced parent sets [BC15].

Algorithm 1: Parent Set Selection

Input : Variable X_i, domain $V \setminus \{X_i\}$
Output: Score cache *closed*

```
Initialize closed and open
while open is not empty do
    Y ← arg max_{Y∈open} h_i(Y)
    for X_j ∈ V \ {X_i} do
        if Y ∪ {X_j} is not in open and closed then
            Calculate h_i(Y ∪ {X_j})
            Add Y ∪ {X_j} to open
        end
    end
    Calculate g_i(Y)
    Remove Y from open
    Add Y to closed
end
```

2.2.4 Greedy Selection

A faster approach is to perform a greedy search that selects the best node to expand at each iteration [CH92]. This is accomplished by setting $h_i = g_i = sc_i$ as the local score function, and defining *open* and *closed* as before. The algorithm can in fact be made simpler as it is unnecessary to store the values of g_i for explored nodes. The time efficiency of this method often comes at a decreased quality of the parent sets found.

2.2.5 Independence Selection

Independence selection attempts at improving the quality of the parent sets without compromising much the time efficiency [SdCCZ15]. The technique can be seen as an A^* search with an inadmissible heuristic. The heuristic is the BIC_i^* defined as:

$$\mathrm{BIC}_i^*(\mathbf{Y}, \mathbf{Z}) = \mathrm{BIC}_i(\mathbf{Y}) + \mathrm{BIC}_i(\mathbf{Z}) + \mathrm{inter}_i(\mathbf{Y}, \mathbf{Z}), \tag{2.4}$$

where $\mathbf{Y}$ and $\mathbf{Z}$ are two non-empty disjoint sets of variables and $\mathrm{inter}_i(\mathbf{Y}, \mathbf{Z}) = \frac{\log N}{2}(r_i - 1)(\prod_{X_j \in \mathbf{Y}} r_j + \prod_{X_j \in \mathbf{Z}} r_j - \prod_{X_j \in \mathbf{Y} \cup \mathbf{Z}} r_j - 1) - \mathrm{BIC}_i(\emptyset)$. This approximate scoring function can be calculated in constant time (i.e., $O(1)$ complexity) if $\mathrm{BIC}_i(\mathbf{Y})$ and $\mathrm{BIC}_i(\mathbf{Z})$ are cached. Additionally, the local score BIC_i can also be efficiently computed from the BIC_i^* scores as

$$\mathrm{BIC}_i(\mathbf{Y} \cup \mathbf{Z}) = \mathrm{BIC}_i^*(\mathbf{Y}, \mathbf{Z}) + N \cdot I_i(\mathbf{Y}, \mathbf{Z}), \tag{2.5}$$

where I_i is the *Interaction Information* estimated from data [McG54].

The approach uses $h_i(\mathbf{Y} \cup \mathbf{Z}) = \mathrm{BIC}_i^*(\mathbf{Y}, \mathbf{Z})$ and $g_i(\mathbf{Y}) = \mathrm{BIC}_i(\mathbf{Y})$. The list *closed* is initialized containing all the parent sets with cardinality at most one (including the empty set), and *open* is initialized with all the parent sets of cardinality two.

Since the scores are calculated for increasingly larger parent set sizes, pruning rules can be used to detect unnecessary computations; also the BIC_i^* are always available from cached scores at each iteration, and BIC_i score values can be obtained efficiently by computing $I_i(\mathbf{Y}, \mathbf{Z})$. However, using some pruning rules can make some possibly better BIC_i^* values impossible to be calculated using equation 2.4 because the corresponding BIC_i scores are not cached on memory.

Row	X	Y	Z
1	x_1	y_1	z_1
2	x_1	y_1	z_2
3	x_1	y_2	z_1
4	x_2	y_2	z_2
5	x_2	y_1	z_1
6	x_2	y_1	z_2
7	x_2	y_2	z_1
8	x_3	y_2	z_2
9	x_3	y_1	z_1

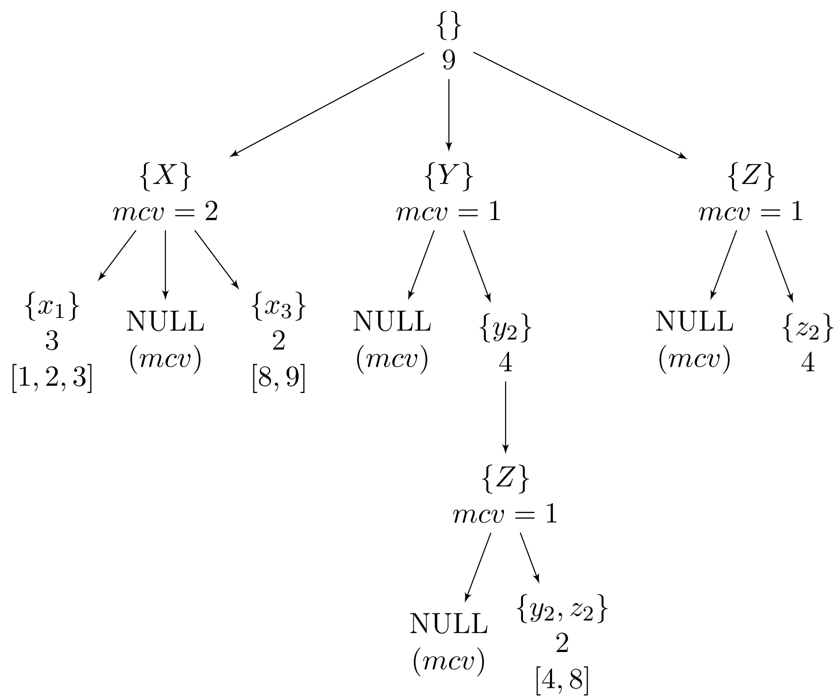


Figure 2.4: *AD-Tree for a domain X, Y, Z.*

2.2.6 AD-Tree

There is a second issue related to counting statistics from the data set given a variable instantiation, since they are necessary for calculating scores. An AD-tree is an unbalanced tree used to store statistics from a data set that allows to perform queries over instantiations of variables in an efficient way [ML98]. It has two types of nodes: AD-nodes and Varying nodes which store number of records consistent with the variable instantiation of the node and assign a new variable, respectively.

A *full* AD-tree have R^n nodes when each variable of the domain has cardinality R. An *sparse* AD-tree is built with lesser nodes by performing some optimizations which help to reduce memory consumption and speed counting.

Some of the optimizations are to omit AD-nodes with zero instances and to stop expanding Varying nodes having the greatest count, called *most common value* (MCV). Moreover, a global parameter $rMin$ is used to control the sparseness of the structure, that is, AD-nodes with less than $rMin$ instances are not expanded. By performing these optimizations the structure can be built in worst-case time $O(nN + n^{\log_R N/rMin}/\log_R N)$.

For instance, a data set and its corresponding AD-tree built using $rMin = 4$ is showed on Figure 2.4. Its root is an AD-node with the total number of records. Its children are the varying nodes $\{X\}$, $\{Y\}$ and $\{Z\}$ each having an attribute mcv representing its child with the greatest counter marked as a $NULL$ node. Note that AD-node $\{y_2\}$ is expanded because its counter is not less than $rMin$, but other AD-nodes have a list with index instances holding its instantiation.

Once the AD-tree was built, a query can be answered in worst-case time $O(sN + sR^{s-1})$ where s is the number of variables in the query, amortized time $O(\min(sN, sR^{s-1}))$.

2.3 Order-based structure optimization

As discussed in Section 2.1, another important step in order-based learning consists in obtaining the network G^* which maximizes the score $sc(G^*)$. Although, there are some exact methods that guarantee obtaining the best network given all $\mathbf{C}_i$, they are computationally costly [PIM08, YMW11, YM12, BC15]. In this section, we describe the state-of-the-art order-based local search methods which are the most competitive approximate approaches for structure learning on large domains.

2.3.1 Greedy Search

Greedy search is an efficient method that remains as the state-of-the-art of structure learning approximated methods. The method relies on obtaining a DAG G, given an ordering L, by the following equation:

$$sc(L) = sc(G) = \sum_{i=1}^{n} \max_{\mathbf{Y} \in \mathbf{C}_i^L} sc_i(\mathbf{Y}). \tag{2.6}$$

where $\mathbf{C}_i^L = \{s \in \mathbf{C}_i \mid \forall X_j \in s, X_j < X_i\}$, denotes the parent sets in $\mathbf{C}_i$ containing variables X_j that appear before it in the order L. For instance, Figure 2.1 shows a DAG G consistent with the ordering $[A, B, C, D]$.

The algorithm, shown on Algorithm 2, starts with an initial order (Line 1) as the incumbent solution. Then, for a determined number of iterations K, it explores the incumbent solution's neighborhood and get its best neighbor (i.e., the one with greater score) on Line 3. The neighborhood of an ordering L is the set of orderings with two adjacent variables X_i, X_{i+1} swapped (i.e., exchange positions). If the best neighbor improves the incumbent solution's score, the algorithm replace the incumbent solution with the it (Line 5). Other case, the search stops because it reaches a local optimum (Line 7). Finally, it returns a DAG G consistent with the incumbent solution L (Line 10).

Algorithm 2: Greedy search

```
   Input  : Candidate sets C_i with their scores pre-computed
   Output: A DAG G
 1 Initialize L
 2 for j = 1 to K do
 3 |   bestNeighbor ← getBestNeighbor(L)
 4 |   if bestNeighbor is better than L then
 5 |   |   L ← bestNeighbor
 6 |   else
 7 |   |   Stop the search
 8 |   end
 9 end
10 return DAG G consistent with L
```

The main advantage of this approach is that when swapping two adjacent variables, to update the score is computationally efficient. For example, by swapping variables X_i and X_{i+1} on the ordering L, we obtain a new ordering L' such that

$$sc(L') = sc(L) - sc_i(Pa_i) - sc_{i+1}(Pa_{i+1}) + sc_i(Pa'_i) + sc_{i+1}(Pa'_{i+1}), \tag{2.7}$$

where Pa'_i is the parent set of the variable X_i on DAG G' consistent with L'.

2.3.2 Simulated Annealing

The problem with previous method is that it does not allow to visit neighbors that do not improve the incumbent solution, therefore it can lead to a poor local maxima. Simulated annealing tries to escape poor local maxima by visiting the incumbent solution's neighborhood with a probabilistic approach [GKR94].

The algorithm, shown on Algorithm 3, uses Equation (2.6) and (2.7) to obtain and update the score of an ordering, respectively. It starts with an initial ordering (Line 1) and a temperature parameter T (Line 2), usually set to high values. Then, for a determined number of iterations

K, it gets an incumbent solution's neighbor uniformly at random (Line 4), hence not necessarily the one with greater score. If the random neighbor improves the incumbent solution's score, the algorithm perform a movement to it (Line 6). Other case, the method replaces the incumbent solution with the poor-quality random neighbor with probability $exp(\frac{sc(L)-sc(neighbor)}{T})$ (Line 10). After each iteration the temperature T is reduced in order to decrease the probabilities to move to worse solutions. Finally, it returns a DAG G consistent with the incumbent solution L (Line 15).

Algorithm 3: Simulated Annealing

```
   Input  : Candidate sets C_i with their scores pre-computed
   Output: A DAG G
 1 Initialize L
 2 Initialize temperature T
 3 for j = 1 to K do
 4     randomNeighbor ← getRandomNeighbor(L)
 5     if randomNeighbor is better than L then
 6         L ← randomNeighbor
 7     else
 8         P ← exp((sc(L) − sc(randomNeighbor)) / T)
 9         if Accept with probability P then
10             L ← randomNeighbor
11         end
12     end
13     Reduce temperature T
14 end
15 return DAG G consistent with L
```

The main disadvantage of this approach is that visiting the neighborhood of an ordering uniformly at random hurts the performance of the method, therefore it is needed to perform several re-starts (i.e., run multiple times).

2.3.3 Tabu Search

The problem with greedy search is there are swaps which do not improve the incumbent solution's score and only increase the calculation time needed. Such swaps, referred here as *tabu swaps*, can be omitted on following iterations. The search uses a *tabu list* to save the last S tabu swaps on the search [Glo89].

The algorithm, shown on Algorithm 4, also uses Equation (2.6) and (2.7) as previous methods. It starts with an initial ordering (Line 1) as the incumbent solution and an empty *tabu list* (Line 2). Then, for a determined number of iterations K, it explores the incumbent solution's neighborhood ignoring the neighbors obtained by the last S tabu swaps (Line 4). While visiting the neighborhood of the incumbent solution, the tabu list is updated with the last S tabu swaps. After that, the search behaves as greedy search, always accepting better orderings (Line 6) and stopping when a local optimum is reached (Line 8). Finally, it returns a DAG G consistent with the incumbent solution L (Line 11).

The main advantage of this approach is that by maintaining a tabu list, it is not necessary to calculate the score of certain neighbors. However, in some cases, adopting a considerably high value of S, hurts the performance of the method.

2.3.4 Beam Search

The disadvantage of previous approaches is that they only visit the neighborhood of a single incumbent solution, possibly leading to a poor local maxima. However, there is always a possibility

Algorithm 4: Tabu search

```
Input  : Candidate sets C_i with their scores pre-computed
Output: A DAG G
Initialize L
Initialize tabu list T as empty
for j = 1 to K do
    bestNeighbor ← getBestNeighbor(L, T)
    if bestNeighbor is better than L then
        L ← bestNeighbor
    else
        Stop the search
    end
end
return DAG G consistent with L
```

that other neighbors' neighborhood leads to better results. In this manner, Beam search extends the idea of greedy search by keeping in memory multiple incumbent solutions. The later increases the possibilities of escaping poor local maxima, but also adds a memory overhead.

Algorithm 5: Beam search

```
    Input  : Candidate sets C_i with their scores pre-computed
    Output: A DAG G
 1  Initialize L
 2  Initialize Q with L
 3
 4  for j = 1 to K do
 5      Q ← getqBestNeighbors(Q, q)
 6      bestNeighbor ← arg max_< Q
 7      if bestNeighbor is better than L then
 8          L ← bestNeighbor
 9      end
10  end
11  return DAG G consistent with L
```

The definitions on Equation (2.6) and (2.7) also applies to this approach. The algorithm, shown on Algorithm 5, starts with an initial ordering (Line 1) and an empty list of the last q incumbent solutions Q (Line 2). Then, for a determined number of iterations K, it visits each incumbent solution's neighborhood and update Q with the best q orderings of all neighbors (Line 5). After that, the best solution L is updated if the best ordering on Q has a greater score (Line 8). Finally, it returns a DAG G consistent with the best ordering L (Line 11).

Beam search differs from multiple runs of greedy search in that the intermediate solutions exchange information (in the sense that only the q best neighbors of all runs are kept from one iteration to another). The main disadvantage of this approach relies on its memory overhead, that is why q is usually set to lower values (e.g., $\log n$) for large domains.

2.3.5 Acyclic selection order-based search

Although using Equation 2.7 is efficient for updating the score by refusing all cyclic structures, this also prunes some acyclic structures because of the way of interpreting the order. Acyclic selection tries to improve this by not only considering candidate parent sets with variables before X_i, but also parent sets that do not introduce a cycle in the *current* partial network [SdCCZ15].

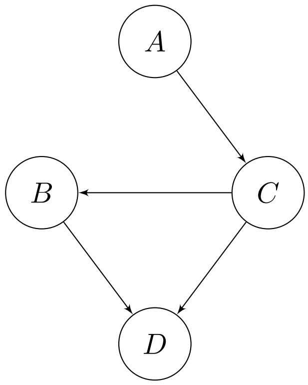


Figure 2.5: *DAG G acyclicly consistent with order* $[A, B, C, D]$

To check the acyclicity of the graph at every step has high computational cost. However, this approach can be implemented in linear time, thus matching the asymptotic performance of the greedy approach [SdCCZ15].

Given an ordering $X_1 < X_2 < \ldots < X_n$, the algorithm starts with an empty G_n and a Boolean matrix M representing if variable X_j is descendant of variable X_i. Then, for $i = n$ to $i = 1$, acyclic selection searches for the best parent set for X_i that does not induce cycles in G_{i-1}, and sets G_i to be G_{i-1} with the selected parents of X_i. After that, it saves the descendants of X_i in a list in order to perform a depth-first search on the incumbent network and update the descendants of each variable on matrix M. Although performing a depth-first search could increase the complexity, the algorithm remains linear because it does not need to visit nodes that were not changed.

For instance, consider the same order from the greedy approach's example, Figure 2.5 shows an *acyclicly* consistent DAG G with the order. Notice that C is parent of B, but it appears after B because doing that does not introduce a cycle and it is a better parent set than $\{A\}$.

Scanagatta et. al proves that for a given ordering L, the DAG G obtained by using an acyclic selection will get results equal or greater than the one obtained by the greedy search [SdCCZ15]. Then, by combining acyclic selection with greedy search, they get a novel approach, known as ASOBS, which obtains better results than greedy search.

A disadvantage of ASOBS is that there is not an efficient way yet to recalculate the score when swapping adjacent variables. Moreover, the algorithm has a memory overhead based on the necessity of the auxiliar matrix M.

2.3.6 Swap Search

Greedy search is extremely efficient due to the constant time evaluation when moving to a new ordering on the neighborhood of ordering L. On the other hand, Acyclic selection finds a better DAG for the same ordering L, but taking a slightly greater complexity time. We here propose a hybrid approach that takes advantage of both methods.

The hybrid approach, which we called *Swap search*, combines the two state-of-the-art structure learning methods is shown on Algorithm 6. The search starts using a greedy search to obtain DAGs given an ordering, but when the incumbent solution converges, it swaps to the ASOBS algorithm, and viceversa. This allows to combine the efficient swap of the greedy approach with the great performance obtained with acyclic selection.

Algorithm 6: Swap search

```
Input  : Candidate sets C_i with their scores pre-computed
Output: A DAG G
Initialize L
for j = 1 to K do
    bestNeighbor ← getBestNeighbor(L)
    if bestNeighbor is better than L then
        L ← bestNeighbor
    else
        if Is in greedy search then
            Change score calculation to acyclic selection
        else
            Change score calculation to greedy search
        end
    end
end
return DAG G consistent with L
```

Chapter 3

Contributions

3.1 Generating informed initial solutions

As with most local search approaches, the selection of a good initial solution is crucial for avoiding convergence to poor local maxima in order-based structure optimization. Typically, this is attempted by randomly generating initial orderings using standard techniques such as the Fisher-Yates algorithm [Knu98]. While this guarantees a good coverage of the search space when sufficiently many re-starts are performed, in large domains it can lead to poor solutions and require many iterations until a local optimum is reached. In this chapter, we devise heuristics that take advantage of the formulation of the problem to produce better initial solutions.

3.2 Best Parent Set Graph

A parent set Pa_i that maximizes the score function for a variable X_i is called the *best parent set*. In a similar manner, a graph H^* is called the *best parent set graph* if $Pa_i^{H^*}$ satisfies

$$Pa_i^{H^*} \in \arg\max_{\mathbf{Y}} sc_i(\mathbf{Y}) \tag{3.1}$$

Note that, this graph usually contains cycles because it has the best parent set for every variable without considering a prefixed order; thus, it is not a solution to Equation (2.2), but an upper bound on the value of $sc(G^*)$.

The best parent set graph can be used to reduce the search space of the problem in (2.2), by finding $\tilde{G}$ such that

$$sc(\tilde{G}) = \max_{<} \sum_{i=1}^{n} \max_{\mathbf{Y} \in \mathbf{C}_i^L \cap Pa_i^{H^*}} sc_i(\mathbf{Y})\,. \tag{3.2}$$

In addition, the best parent set graph H^* can be used as the base of initialization heuristics for generating topological orders which will be covered in next subsections. In practice, obtaining the graph H^* can be difficult, and we often resort to an approximate solution H where the parents Pa_i^H are obtained using one of the parent set selection methods described in Section 2.2.

3.3 DFS-based approach

We can exploit the information provided by the best parent set graph on Equation (3.1) or an approximation of it to bias the generation of topological orderings towards high-scoring regions. To see this, consider a(n approximation of the) best parent set graph with nodes X_i and X_j such that X_i is the single parent of X_j and has no parents. Then, there is an optimal ordering starting with X_i (this can easily be shown by contradiction). We can delete X_i from the graph and repeat the argument to conclude the existence of an optimal ordering starting with X_i, X_j. Now consider a case when there are two or more selectable nodes (by the previous explanation) in graph H. Instead

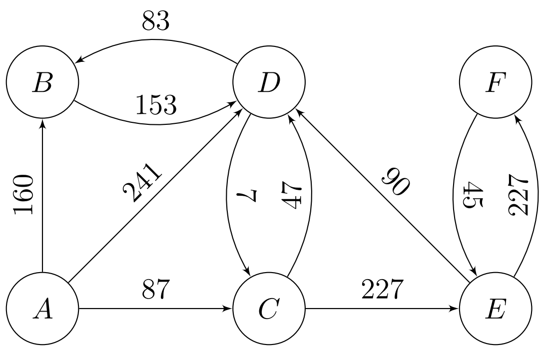


Figure 3.1: *An example of a best parent set graph.*

of picking a random selectable node we can define the *goodness* of a node by:

$$goodness(X_i) = \prod_{X_j \in Ch_i^H \cap unvisited} |Pa_j^H \cap unvisited| \tag{3.3}$$

where Ch_i^H is the set of X_i's children and *unvisited* the set of unvisited nodes. Small values of *goodness* mean that removing X_i from the graph will make more nodes to be selectable. Ties are resolved by picking one of the best selectable nodes uniformly at random.

For example, in the (best parent set) graph in Figure 3.1, we can safely constrain the orderings to start with A, since it has no parents, and remove it from the graph. At this time, we have three selectable nodes B, C and F, each one with same in-degree, but with different goodness value. Since F has the least goodness value, we select it. Performing previous steps repeatedly we get that the candidate optimal orderings are A, F, C, E, B, D and A, F, C, E, D, B. Note that this is a significant decrease from the full space of $6! = 720$ possible orderings. This difference is likely to increase as the number of variables increases, and as the best parent set becomes sparser (the sparsity of the best parent set is related to the score function used, and the ratio between the domain dimension and the data set size).

Motivated by the previous argument, we propose the *DFS-Based initialization heuristic* described in Algorithm 7 which takes a (parent set graph) H as input.

Algorithm 7: DFS-Based ordering generation.

Function: DFS(Graph H)
unvisited $\leftarrow$ all nodes
$L \leftarrow \emptyset$
while *unvisited is not empty* **do**
    $O \leftarrow$ unvisited nodes ordered by unvisited in-degree and goodness
    $B \leftarrow$ best nodes from O
    if B *has more than one node* **then**
        select a node X_r from B uniformly at random
    else
        select the unique node X_r from B
    end
    $L \leftarrow L \cup \{X_r\}$
    $unvisited \leftarrow unvisited \setminus \{X_r\}$
end
return L

The algorithm starts with all nodes labeled as unvisited and repeatedly selects an unvisited node using as criterias the in-degree and goodness of the node in increasing order. At each step, it marks the node as visited and consider it as removed from H. The algorithm returns the ordering when all nodes were visited. Table 3.1 shows a possible execution of the algorithm which only selects a node uniformly at random in iteration 5.

Iteration	Metric	A	B	C	D	E	F	Node selected
1	in-degree	**0**	2	2	4	2	1	A
	goodness	**3**	3	3	1	0	1	
2	in-degree		**1**	**1**	4	2	**1**	F
	goodness		2	2	0	0	**1**	
3	in-degree		**1**	**1**	3	**1**		C
	goodness		2	**0**	0	2		
4	in-degree		1		2	**0**		E
	goodness		1		0	**0**		
5	in-degree		**1**		**1**			B (at random)
	goodness		**0**		**0**			
6	in-degree				**0**			D
	goodness				**0**			

Table 3.1: *Possible execution of DFS-based generation*

3.4 FAS-based approach

The DFS-based approach can be seen as removing edges from a graph H so as to make it a DAG (more specifically, a tree), and then extracting a consistent topological ordering. The selection of an edge to remove is often performed locally (among the children of a node), and considers only the qualitative information of the graph (i.e., the parent relationships). An arguably better approach is to use the score function to assess the relevance of each edge, and to consider the removal of edges globally (not only in a local neighborhood). We estimate the relevance of an edge $X_j \to X_i$ in a graph H by

$$W_{ji} = sc_i(Pa_i^H) - sc_i(Pa_i^H \setminus \{X_j\}), \tag{3.4}$$

The weight W_{ji} represents the cost of removing X_j from the set Pa_i^H, and it is always a positive number if H is the best parent set graph since Pa_i^H maximizes the score for X_i. A small value of W_{ji} suggests that the parent X_j is not very relevant to X_i. For instance, in the weighted graph in Figure 3.1, the edge $C \to D$ is less relevant than the edge $B \to D$, which in turn is less relevant than the edge $A \to D$.

The main idea of our second heuristic is to penalize orderings which violate an edge $X_i \to X_j$ in H by their associated cost W_{ij}. We then wish to find a topological ordering of H that violates the least cost of edges. Given a directed graph $H = (V, E)$, a set $F \subseteq E$ is called a Feedback Arc Set (FAS) if every (directed) cycle of H contains at least one edge in F. In other words, F is an edge set that if removed makes the graph H acyclic [DF03]. If we assume that the cost of an ordering of H is the sum of the weights of the violated (or removed) edges, we can formulate the problem of finding a minimum cost ordering of H as a Minimum Cost Feedback Arc Set Problem (min-cost FAS): given the weighted directed graph H with weights W_{ij}, find a min-cost FAS F such that

$$F = \arg \min_{H-F \text{ is a DAG}} \sum_{X_i \to X_j \in E} W_{ij}. \tag{3.5}$$

The min-cost FAS problem have been proved to be NP-complete for directed graphs [Gav77], but there are efficient and effective approximation algorithms [EXS93, EL95, DF03] like the one shown in Algorithm 8 with complexity $O(nm)$, where m is the number of edges on the graph.

Algorithm 8: Minimum Cost FAS approximation

```
Input  : Graph H
Output: Feedback Arc Set F
F ← ∅
while there is a cycle C on H do
    W_min ← arg min_{(u,v)∈C} W_uv
    for (u, v) ∈ C do
        W_uv = W_uv − W_min
        if W_uv = 0 then
            F = F + {(u, v)}
        end
    end
end
for (u, v) ∈ F do
    if (u, v) does not introduce a cycle then
        H = H + (u, v)
        F = F \ {(u, v)}
    end
end
```

We can now describe our second heuristic for generating initial solutions, based on the min-cost FAS problem: take the weighted graph H with weights W_{ij} as input, and find a min-cost FAS F; remove the edges in F from H and return a topological order of the DAG $H - F$ (this can be done by performing a depth-first search traversal starting at root nodes).

3.5 Best-first approach

Whereas the DFS- and FAS-based heuristics provide a significant improvement on the quality of solutions found by greedy search, the old state-of-the-art, in fixed amount time, they generate only a marginal gain in performance when ASOBS, the new state-of-the-art, is adopted [PM16]. One possible explanation is that ASOBS performs parent set selection under a (dynamically chosen) variable ordering, hence biasing the search towards specific orderings can actually hurt performance.

Motivated by the previous explanation, we propose the *Best First-Based initialization heuristic* (BFT) described in Algorithm 9. The algorithm takes a collection of possible candidate parent sets $\mathbf{C}_i$ for each variable. The heuristic initially labels all nodes as *unvisited* (Line 1), and initialize the ordering L as empty (Line 2). Then the loop in Lines 3 to 12 generates an ordering as follows. In Line 4, the best valid parent set (i.e., it does not contain visited nodes and have the best score) is selected for each non-visited variable, and then ranked by their score in decreasing order (Line 5). Then a variable is generated with probability proportional to its ranking (Lines 6 to 9). After that, the ordering and the set of visited nodes are updated (Lines 10 and 11, respectively).

Algorithm 9: BestFirst-Based ordering generation.

Function: BestFirst(Candidate Parent Sets C_i)

1 $visited \leftarrow \emptyset$
2 $L \leftarrow \emptyset$
3 **for** $r = n$ *to* 1 **do**
4     $S \leftarrow \{(X_i, bestScore_i^{visited}) \mid X_i \notin visited\}$
5   Sort S decreasingly by $bestScore_i^{visited}$
6   **for** $j = 1$ *to* $|S|$ **do**
7         $Prob_j = \frac{1}{j}$
8   **end**
9     $X_r \leftarrow$ random variable using probability distribution $Prob_r$
10     $L[r] \leftarrow X_r$
11     $visited = visited \cup \{X_r\}$
12 **end**
13 **return** L

The time complexity of the procedure is dominated by the selection of the best parent set for each variable in Line 4. Assuming we uses the efficient implementation of $bestScore_i^{visited}$ developed by Malone based on bitsets [Mal12] , then the complexity of the procedure is as follows. Line 4 is performed in worst-time $\mathcal{O}(nk)$, where k is the maximum cardinality of parent sets. Line 5 sorts the list in time $\mathcal{O}(n \log n)$. The loop from line 6 to 8 is $\mathcal{O}(n)$ and lines 9 to 11 are performed in constant time. Since it is necessary to repeat all steps at each iteration, the overall complexity is $\mathcal{O}(n^2k)$. Notice that if we save all $\mathbf{C}_i$ as lists, the complexity of Line 4 will be $\mathcal{O}(Ck)$, where C is the total number of candidate parent sets. Consequently, the overall complexity will be increased to $\mathcal{O}(Cnk)$, a very inefficient scenario for large domains.

Chapter 4

Results and Discussion

4.1 Experimental Setup

Our experiments aim at evaluating the performance of order-based local search with different parent set selection procedures and different initialization heuristics on a selected set of real-world data sets listed in Table 4.1.[1] The algorithms were implemented in C++, using few utilities from the URLearning package for learning Bayesian networks.[2] All experiments were performed in a 20-node computer cluster; each computer has an Intel Xeon CPU 2.40GHz processor and 512 GB RAM.

For each data set we run sequential, greedy and independence selection with no limit on the maximum parent set size k, but using a time limit of two minutes per variable. After that, we performed 100 re-starts of each order-based local search, each taking at most 500 iterations ($K = 500$) and maximum time of 20 minutes, using the candidate parent sets obtained from the previous step. Appendix A shows a more detailed explanation of each local search algorithm's configuration. For the sake of legibility, we report the quality of a DAG found by its *relative score* given by $RC(G) = \frac{sc(G)-sc(\emptyset)}{|sc(\emptyset)|}$, where $sc(\emptyset)$ is the score of an empty DAG.[3]

We refer to the variant of the order-based learning algorithm with random initialization, depth-first based initialization, min-cost FAS-based and best-first based initialization as RND, DFS, FAS and BFT, respectively.

4.2 Parent set selection approaches

We evaluate the effect the different initialization heuristics have on the quality of the solutions generated by order-based search for the three parent set selection techniques described. Tables 4.1 and 4.2 contain relevant statistics about the parent sets selected for each method. The columns Nps_x and M_x represent, respectively, the number of candidate parent sets and maximum in-degree in the best parent set graph using method x. We see from these results that the maximum degree using greedy or independence selection is, in average, greater than sequential selection, while the amount of candidate parent sets is considerably lower. Some exceptions are the data sets Nltcs and Msnbc where M using greedy selection is the lowest from all procedures probably caused by their small number of variables.

The terms H^*_{seq}, H^*_{gre} and H^*_{ind} in Table 4.2 represents the best parent set graphs obtained for the sequential, greedy and independence selection, respectively, while the columns $D(H^*)$ report the average number of parents in the graph H^*. It is easy to notice that as the number of variables increases, the differences between D also does. Moreover, in average, graphs H^* have not only higher score values when used greedy and independence selection, but also higher average in-degree.

[1] Same data sets used in [SdCCZ15]

[2] The code is available at https://github.com/NonWhite/BN_Solver

[3] A directed acyclic graph where each node has no parents

Dataset	n	N	Nps_{seq}	M_{seq}	Nps_{gre}	M_{gre}	Nps_{ind}	M_{ind}
Nltcs	16	21574	28.8K	6	8.2K	5	28.8K	6
Msnbc	17	388434	233.7K	9	14.1K	6	105.7K	11
Kdd	64	234954	749.6K	4	214.6K	5	21.6K	6
Plants	69	23215	27.4M	4	2.2M	5	57.3K	7
Baudio	100	20000	19.6M	4	4.3M	6	59.5K	6
Bnetflix	100	20000	8.9M	4	2.7M	6	43.9K	6
Jester	100	14116	36.4M	4	5.5M	5	87.9K	5
Accidents	111	17009	2.6M	4	823.8K	7	18.5K	6
Tretail	135	29387	7.5K	4	6.8K	7	5.6K	4
Pumsb_star	163	16349	33.6M	4	1.0M	5	64.6K	7
Dna	180	3186	31.6K	3	26.6K	4	6.0K	4
Kosarek	190	44500	10.9M	3	1.2M	6	100.5K	6
Msweb	294	37711	95.5K	3	71.4K	6	13.2K	8
Book	500	11598	16.3M	3	2.4M	4	388.8K	4
Tmovie	500	6117	19.0M	3	5.8M	5	325.2K	5
Cwebkb	839	4199	8.0M	2	2.4M	5	458.9K	5
Cr52	889	9100	8.2M	2	2.9M	5	411.2K	6
C20ng	910	18821	53.2M	2	10.3M	5	839.7K	5
Bbc	1058	2225	1.1M	2	995.6K	4	306.5K	4
Ad	1556	3279	116.6K	2	136.5K	4	71.4K	4
Average			**12.3M**	**3.6**	**2.1M**	**4.75**	**170.7K**	**5.75**

Table 4.1: *Data sets characteristics*

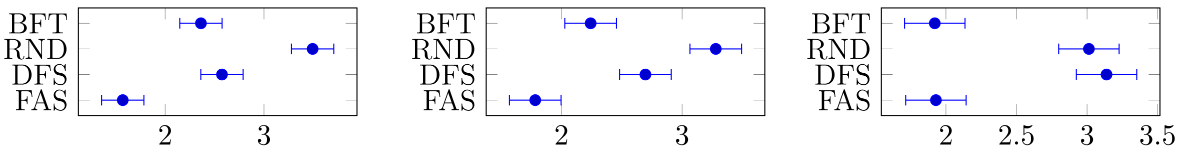
Dataset	$sc(H^*_{seq})$	$D(H^*_{seq})$	$sc(H^*_{gre})$	$D(H^*_{gre})$	$sc(H^*_{ind})$	$D(H^*_{ind})$
Nltcs	0.4515	4.750	0.4359	3.812	0.4515	4.750
Msnbc	0.1660	8.412	0.0867	4.941	0.1755	9.235
Kdd	0.1806	3.438	0.1759	3.672	0.1833	4.078
Plants	0.6839	3.913	0.6685	3.971	0.6631	4.638
Baudio	0.1880	3.590	0.1914	4.520	0.1996	4.980
Bnetflix	0.1336	3.550	0.1343	4.420	0.1441	4.970
Jester	0.1676	3.680	0.1690	4.490	0.1739	4.880
Accidents	0.5797	3.171	0.6314	3.847	0.5352	3.775
Tretail	0.0766	2.111	0.0793	2.081	0.0766	2.096
Pumsb_star	0.8199	2.350	0.7915	2.540	0.7678	2.577
Dna	0.4062	2.856	0.4004	2.806	0.4005	2.617
Kosarek	0.2303	2.668	0.2332	2.979	0.2379	3.342
Msweb	0.1852	1.701	0.1870	1.789	0.1923	2.276
Book	0.1437	2.118	0.1513	2.790	0.1527	2.894
Tmovie	0.3249	2.070	0.3610	2.956	0.3658	3.088
Cwebkb	0.1468	1.895	0.1692	2.777	0.1716	2.939
Cr52	0.1770	1.892	0.2143	2.963	0.2180	3.151
Bbc	0.0865	1.859	0.0979	2.406	0.0992	2.516
Ad	0.8011	1.073	0.8257	1.226	0.8226	1.217

Table 4.2: *Data sets scores*

4.3 Informed initialization heuristics

We first analyze the performance of the different heuristics when each parent set selection method is used. To verify whether the performance differences are statistically significant, we performed *Friedman Test* [Dem06]. In this and all following experiments, we adopt the statistical significance level $\alpha = 0.05$. The computed p-values are 0.00000000000025, 0.00431998334499 and 0.00009479995949 for average initial score obtained using sequential, greedy and independence selection, respectively. Hence, there are statistically significant difference between the scores in all criteria.

We also performed the *post-hoc Nemenyi Test* for the criteria with statistically significant differences in order to decide which pairwise comparisons were significant [Dem06]; this was carried out by calculating the average ranking of each heuristic considering all runs (i.e. all structure optimization algorithms and all data sets). Then, we compute the critical distance $CD = q_\alpha \sqrt{\frac{k(k+1)}{6m}}$, where k is the number of models to compare (i.e. heuristics), m the number of datasets used and q_α, the penalization factor for multiple comparison, was taken from [Dem06]. The results are presented graphically in Figure 4.1. Each point represents the average ranking of the corresponding approach, and the intervals indicate the critical distance. A method A is considered statistically significant better than a method B (w.r.t. to a specific criterion) if A has a smaller average ranking and their intervals do not overlap.



(a) *Sequential sel.* **(b)** *Greedy sel.* **(c)** *Independence sel.*

Figure 4.1: *Visualization of the Nemenyi post-hoc analysis: Average initial score.*

We see from the figure that FAS outperforms all other initialization heuristics using sequential and greedy selection. However, when using independence selection, BFT has no significantly difference compared to FAS. On the other hand, BFT usually leads to better results than DFS using any parent set selection method, but there is only statistically significant difference when using independence selection.

Although independence selection obtains larger candidate parent sets, the results show that it considerably hurts the performance obtained by DFS-based heuristic compared to RND because of the high in-degree of the best parent set graph. Finally, experiments show that the quality of candidate parent sets obtained are important for the performance of our heuristics.

4.4 Structure optimization methods

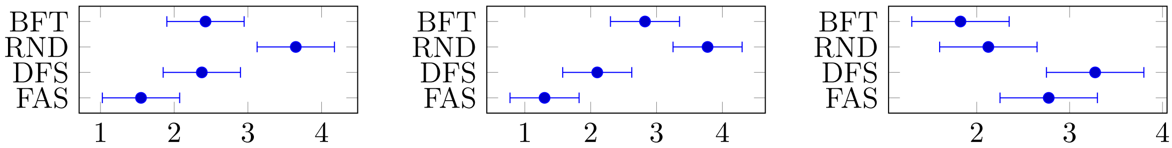
We already know that our heuristics outperform the state-of-the-art random approach to generate initial orders, we can analyze the results obtained by using those generated initial orders on local search methods previously explained in Section 2.3. The complete list of empirical results can be found from Appendix B to G.

Results were compared by the best score and average best score of the Bayesian networks obtained with each data set. Additionally, the average number of iterations necessary to converge (i.e. reach a local maxima) is compared.

4.4.1 Greedy Search

First, we analyze the performance when sequential parent set selection is used. Notice that without the time limit, sequential selection generates the optimal best parent sets. The computed p-values are 0.0000040, 0.0000001 and 0.0010949 for best score, average best score and average number of iterations, respectively. Hence, there are statistically significant difference between the heuristics in all criteria.

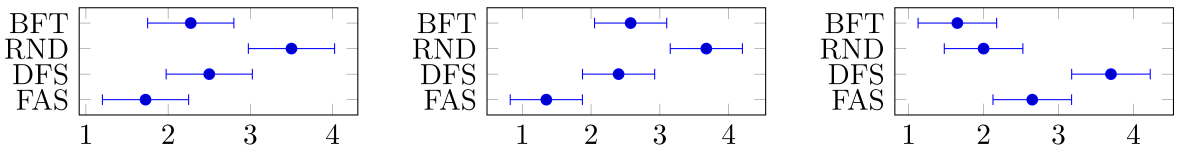
The post-hoc Nemenyi test graphics are shown in Figure 4.2 where we see that the best score and average best score obtained with FAS is significant better compared to RND, but it compares equally with DFS and BFT.



(a) *Best score* **(b)** *Avg. best score* **(c)** *Avg. number of iterations*

Figure 4.2: *Visualization of the Nemenyi post-hoc analysis: greedy search + sequential selection.*

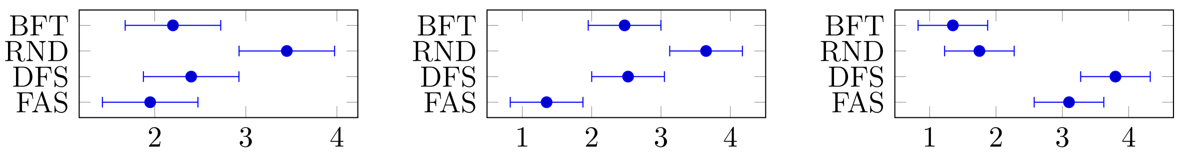
By performing the same statistical analysis when using greedy selection, the obtained p-values are 0.000177, 0.0000001 and 0.000002, respectively for each criterion. In this case, there is also significant difference between the heuristics under any criteria; thus, they are analyzed with the Nemenyi test. Figure 4.3 shows that FAS gets better networks (i.e. with higher scores) than other heuristics, but the difference is not significant. However, on average, FAS outperforms significantly the other heuristics. Although all heuristics performs equally on average number of iterations, BFT and RND need less iterations to converge.



(a) *Best score* **(b)** *Avg. best score* **(c)** *Avg. number of iterations*

Figure 4.3: *Visualization of the Nemenyi post-hoc analysis: greedy search + greedy selection.*

Using independence selection, the Friedman test obtained p-values 0.001331, 0.000001 and 0.0000001, which shows that there are statistically significant differences under any criteria. Experiments show that FAS overall outperforms the other methods on average best score, but there is not guarantee to obtain always best scores than other heuristics. Note that BFT and RND helps greedy search to converge significantly faster than DFS and FAS.



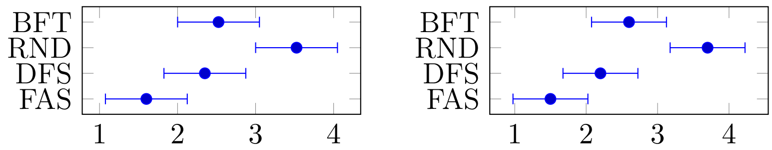
(a) *Best score* **(b)** *Avg. best score* **(c)** *Avg. number of iterations*

Figure 4.4: *Visualization of the Nemenyi post-hoc analysis: greedy search + independence selection.*

Although independence selection makes FAS and greedy search to be significantly better on average, it also hurts considerably the convergence of the search. On the other hand, greedy selection compares equally on average number of iterations and it is also significantly better on average. By the previous fact using FAS and greedy selection is the best setup for greedy search.

4.4.2 Simulated Annealing

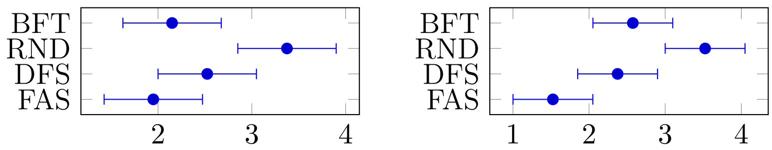
In order to analyze the performance of our heuristics on simulated annealing, we perform the same statistical analysis and obtain p-values 0.000023, 0.0000001 and 0.080594, for best score, average best score and average number of iterations, respectively. Since average number of iterations did not have a p-value < 0.05, it was not analyzed with Nemenyi test. Average rankings and critical distance for each criteria are shown at Figure 4.5. Notice that FAS and DFS get significant better networks that RND, but they are compared equally to BFT under all criteria.



(a) *Best score* **(b)** *Avg. best score*

Figure 4.5: *Visualization of the Nemenyi post-hoc analysis: simulated annealing + sequential selection.*

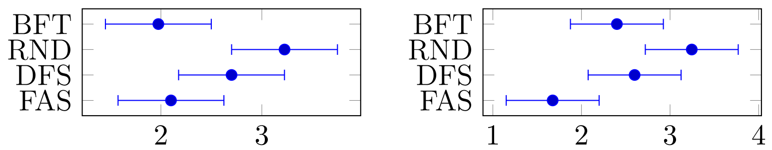
On the other hand, the p-values obtained for greedy selection are 0.001711, 0.000019 and 0.058173 which again shows that there is no statistically significant difference on average number of iterations. In this case, although FAS obtains best networks than DFS and BFT, the test shows that there is no significant difference, except with RND.



(a) *Best score* **(b)** *Avg. best score*

Figure 4.6: *Visualization of the Nemenyi post-hoc analysis: simulated annealing + greedy selection.*

Statistical tests with independence selection shows a similar scenario from before, but also shows that it hurts considerably DFS heuristic, getting closer to RND under all criteria. The p-values obtained by Friedman test are 0.004660, 0.000547 and 0.485243 for each criterion respectively, which shows that heuristics help the algorithm converges equally.



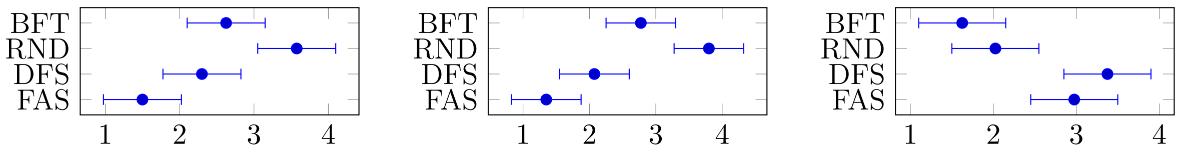
(a) *Best score* **(b)** *Avg. best score*

Figure 4.7: *Visualization of the Nemenyi post-hoc analysis: simulated annealing + independence selection.*

In summary, the best configuration for simulated annealing is to use sequential selection and FAS. Although BFT, DFS and FAS are compared equally, the latter obtains better results.

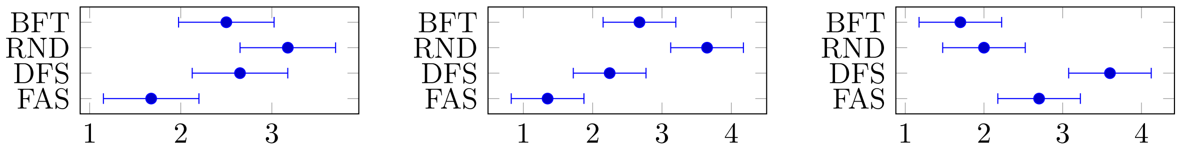
4.4.3 Tabu Search

Tabu search's results using sequential and greedy selection behaves similar to previous algorithms, obtaining p-values 0.000007, 0.0000001 and 0.000015 for sequential, and 0.002940, 0.0000001 and 0.000011 for greedy selection, then making possible the Nemenyi post-hoc test analysis for all criteria.



(a) *Best score* **(b)** *Avg. best score* **(c)** *Avg. iterations*

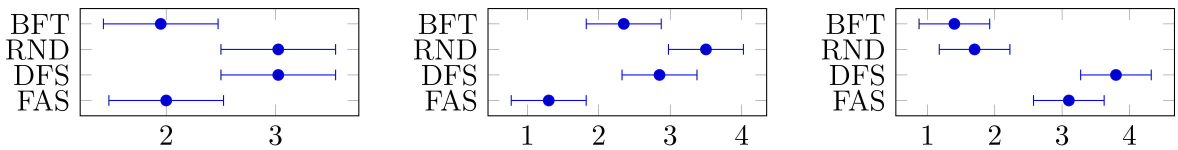
Figure 4.8: *Visualization of the Nemenyi post-hoc analysis: tabu search + sequential selection.*



(a) *Best score* **(b)** *Avg. best score* **(c)** *Avg. iterations*

Figure 4.9: *Visualization of the Nemenyi post-hoc analysis: tabu search + greedy selection.*

However, the use of independence selection creates a very different scenario shown at Figure 4.10, where p-values are 0.003764, 0.000001 and 0.0000001. Graphics show that BFT and FAS obtain statistically significant better networks than RND and DFS. Also, on average, FAS outperforms all heuristics significantly, but it need considerably significant many iterations to converge compared to BFT and RND.



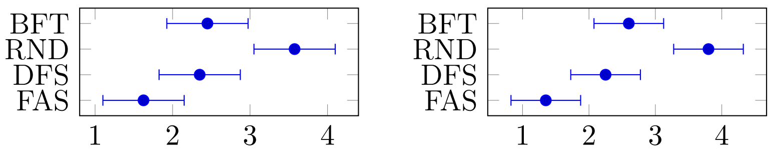
(a) *Best score* **(b)** *Avg. best score* **(c)** *Avg. iterations*

Figure 4.10: *Visualization of the Nemenyi post-hoc analysis: tabu search + independence selection.*

Based on the previous analysis, we get that the best setup for tabu search is to use independence selection and FAS.

4.4.4 Beam Search

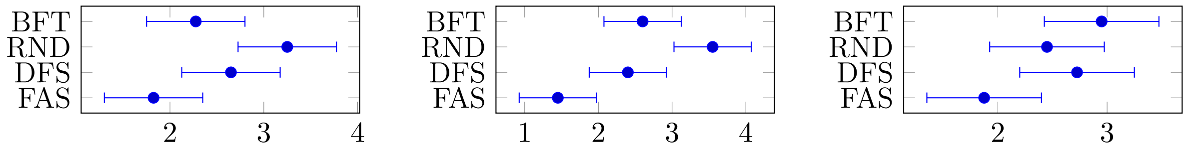
Friedman test's p-values for beam search using sequential selection are 0.00024, 0.0000001 and 0.804315 for best score, average best score and average number of iterations, respectively. Then, all criteria except for average number of iterations were analyzed with Nemenyi test showed in Figure 4.11. FAS, BFT and DFS can be compared equally, but they are significantly better than RND initialization heuristic under all three criteria.



(a) *Best score* **(b)** *Avg. best score*

Figure 4.11: *Visualization of the Nemenyi post-hoc analysis: beam search + sequential selection.*

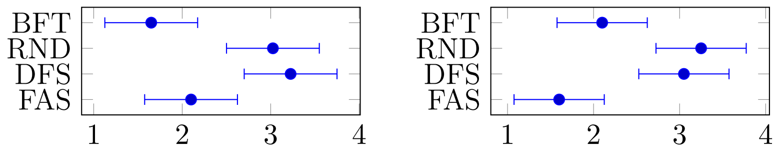
Additionally, p-values for greedy selection are 0.004301, 0.000007 and 0.048709 for each criterion. Figure 4.12 shows that the parent set method considerably hurts the perfomance of our initialization heuristics because of the possible poor quality of parent sets obtained. However, FAS is still significant better to RND in best score and average best score.



(a) *Best score* **(b)** *Avg. best score* **(c)** *Avg. number of iterations*

Figure 4.12: *Visualization of the Nemenyi post-hoc analysis: beam search + greedy selection.*

In contrast, when using independence selection the p-values are 0.000140, 0.000065 and 0.456428 which shows that there is no significant difference on average number of iterations. Note in Figure 4.13 that BFT and FAS outperform RND and DFS under any criteria, but FAS is better on average.



(a) *Best score* **(b)** *Avg. best score*

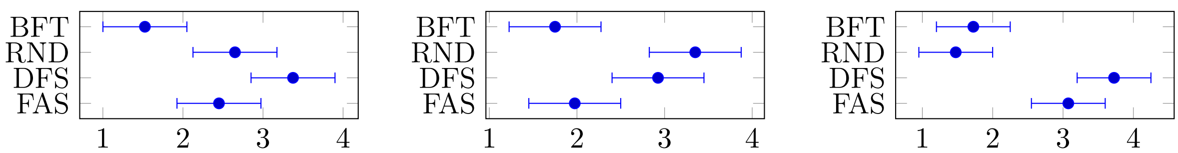
Figure 4.13: *Visualization of the Nemenyi post-hoc analysis: independence search + independence selection.*

Statistical tests show that the best parent set method for beam search is to use independence selection instead of greedy selection because there is no difference in convergence, and of sequential selection because the difference is not remarkably different between heuristics. Also, it is better to use FAS instead of BFT because the former gets better results on average, although they does not have significant difference between them.

4.4.5 Acyclic Selection Order-based Search

The results for the new state-of-the-algorithm for approximated structure learning algorithms shows a considerably different scenario from previous local search methods. Friedman test's p-values for sequential selection are 0.000069, 0.000095 and 0.0000001 for best score, average best score and average number of iterations, respectively. Then, all criteria were evaluated with Nemenyi test and shown in Figure 4.14.

The figure shows that BFT is a very reliable choice for initialize the search when acyclic selection is adopted [PM16]. BFT obtains significant better networks than all other initialization heuristics, but on average its results are compared equally to FAS' ones. Additionally, it needs significant smaller number of iterations to converge, except compared to RND.

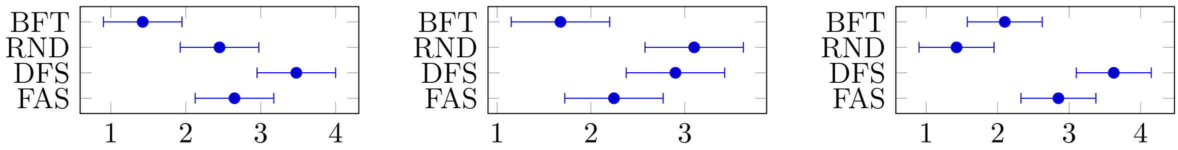


(a) *Best score* **(b)** *Avg. best score* **(c)** *Avg. number of iterations*

Figure 4.14: *Visualization of the Nemenyi post-hoc analysis: asobs search + sequential selection.*

Greedy selection hurts a lot on the performance of the heuristics represented on the p-values 0.000010, 0.001950 and 0.0000001, respectively for each criterion. This impact can be noticed in

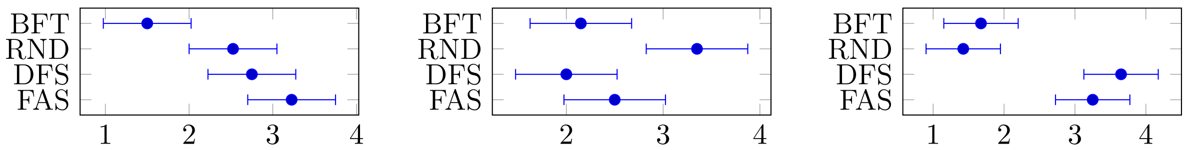
Figure 4.15 where the overlap between heuristics' results are greater than before making almost all compared equally. However, the results behavior remains unchanged (BFT better than others).



(a) *Best score* **(b)** *Avg. best score* **(c)** *Avg. number of iterations*

Figure 4.15: *Visualization of the Nemenyi post-hoc analysis: asobs search + greedy selection.*

ASOBS performance using the great quality list of candidate parent sets from independence selection obtains p-values 0.000210, 0.003711 and 0.0000001 for each criteria, respectively. Since all p-values are smaller than 0.05, the Nemenyi test is performed and showed at Figure 4.16. BFT continues to be the best choice for obtaining good quality Bayesian networks (measured by its score).



(a) *Best score* **(b)** *Avg. best score* **(c)** *Avg. number of iterations*

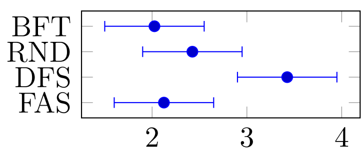
Figure 4.16: *Visualization of the Nemenyi post-hoc analysis: asobs search + independence selection.*

Finally, it is clearly that the best initialization heuristic for ASOBS is BFT. Moreover, the impact of using independence selection is remarkable because it helps to get greater differences between the use of any heuristic. In that manner, the best setup for ASOBS is to use BFT with independence selection.

4.4.6 Swap Search

This new approach takes the advantages not only of greedy search, but also of ASOBS. Swap search obtains p-values 0.858363, 0.477712 and 0.086073 on the Friedman test. Then, none of the criteria were analyzed by the Nemenyi test using parent sets from sequential selection.

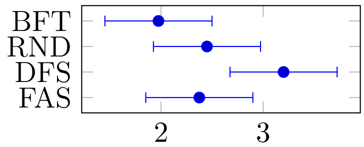
On the other hand, there is a tie between BFT, RND and FAS on the average best score as shown in Figure 4.17 given the p-values 0.109340, 0.001856 and 0.165302, respectively for all criteria. Since RND is the state-of-the-art initialization heuristic, we can state that swap search and greedy selection does not lead to significant better results.



(a) *Avg. best score*

Figure 4.17: *Visualization of the Nemenyi post-hoc analysis: swap search + greedy selection.*

By performing Friedman test when using independence selection we get p-values 0.013541, 0.906527 and 0.165302. Then we perform the Nemenyi test only for the best score field and shown in Figure 4.18. The graphic shows that BFT leads to better Bayesian networks with swap search, but the difference is not considered as significantly different to RND or FAS. It also shows that DFS has a poor performance with swap search compared to other heuristics.



(a) *Best Score*

Figure 4.18: *Visualization of the Nemenyi post-hoc analysis: swap search + independence selection.*

In summary, the performance of solutions obtained is not significantly modified by selecting a parent set selection method. But independence selection get better networks in almost all datasets. Moreover, the best initialization heuristic for swap search is BFT, although the difference is not significant.

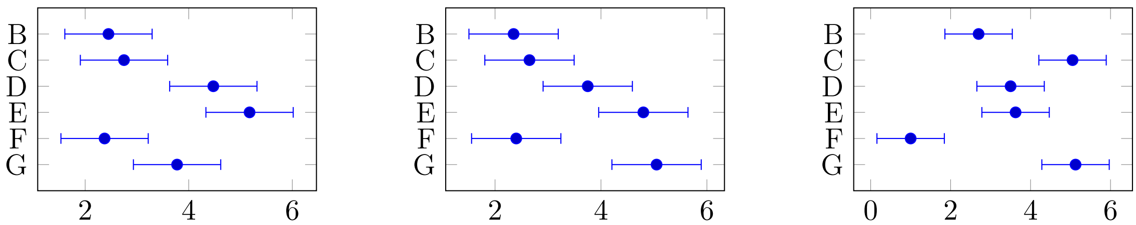
4.5 Discussion

Table 4.3 shows the best setup for each algorithm based on the statistical analysis from previous section.

Appendix	Algorithm	Parent set selection	Initialization heuristic
B	Greedy search	Greedy	FAS
C	Simulated annealing	Sequential	FAS
D	Tabu search	Independence	FAS
E	Beam search	Independence	FAS
F	ASOBS	Independence	BFT
G	Swap search	Independence	BFT

Table 4.3: *Best setup for each algorithm*

Previous table shows that FAS leads to better results when Equation 2.6 is used to calculate the score of an ordering. On the other hand, when using acyclic selection as part of the search, it is always best to choose BFT initialization heuristic.



(a) *Best score* **(b)** *Avg. best score* **(c)** *Avg. number of iterations*

Figure 4.19: *Visualization of the Nemenyi post-hoc analysis: best setups.*

In order to know if there is a significant difference among the best setups, we perform the Friedman test. The p-values obtained were 0.0000002, 0.00000004 and 0.0000000000002 for best score, average best score and average number of iterations, respectively. Then, we perform the Nemenyi test to all criteria and obtain the graphics on Figure 4.19.

The statistical analysis show that greedy search, simulated annealing and ASOBS' setups leads to better significant networks compared to other algorithm's setups. However, ASOBS' setup shows a considerably significant advantage to converge among other setups.

Beam search often leads to results similar to greedy search, but the time limit adopted for each search hurts considerably its performance. Additionally, the performance of tabu search was conditioned by the size of the tabu list which increases the memory overhead and also makes the runs execute in more time. On the other hand, swap search combines greedy search and ASOBS, but it has a high computational cost which makes the seach to exceed the time limit adopted on experiments.

The results show that the scores obtained by two state-of-the-art using greedy search and acyclic selection are compared equally, possibly because of implementation details. Although acyclic selection converges significantly faster than greedy search, the former needs twice the computational time at each iteration. Finally, the best setup for order-based structure learning approximated algorithms depends on the trade-off between efficiency/convergence required.

Chapter 5

Conclussions and Future Work

5.1 Conclusions

Learning Bayesian networks from data is a notably difficult problem, and practitioners often resort to approximate solutions. A state-of-the-art approach for large domains is order-based structure optimization, which performs a local search in the space of variable orderings by avoiding cyclic structures. As with many local search approaches, the quality of the solutions produced by order-based learning strongly depends on the initialization strategy. In this work, we proposed three new informed heuristics for generating initial solutions for order-based structure learning. The first and second heuristics are based on the best parent set graph, which is the directed graph obtained as the solution of the relaxed problem when cycles are permitted. The first heuristic performs a depth-first search traversal of the parent set graph that orders nodes based on their in-degree. The second heuristic uses the score function to assess the relevance of edges in the parent graphs, and finds a minimum weight ordering by solving a minimum-cost feedback arc set problem. Finally, the last heuristic obtains a relaxed version of the problem by allowing getting parent sets containing variables that does not generate a cyclic structure. Note that the last heuristics was developed specifically for the ASOBS method.

Moreover, independence selection shows to be the new state-of-the-art for parent set selection methods because its scalability for working with large domains and the great impact on the performance of the structure optimization methods.

Experiments with 20 real-world data sets containing from 16 to 1556 variables demonstrate that our initialization heuristics improve the accuracy of order-based local search methods. To our knowledge, these are the largest domains considered in the literature.

5.2 Future work

Higher values of scoring functions mean the domain is well represented by the structure of a Bayesian network; that is why often Bayesian networks are used for *traditional classification* which given a feature instantiation, returns a class label by performing inferences on the model. Some of the main methods to build traditional classifiers are Tree-Augmented Tree (TAN), Naive Bayes (NB) and General Bayesian Networks (GBN) [FG96, Mad08, CG99]. Nevertheless, a more generic and complex problem called *multi-label classification*, where instances has observations about features and multiple class labels, extends the problem to return a vector of class labels given a feature instantiation. Some state-of-the-art approaches tries to maintain a trade off between speed and accuracy of classifiers learned, but they make some assumptions on the relationships among variables and classes that may hurt the performance of classifiers and in some cases allowing them to avoid structure learning [ACMG13, dWG07]. A proposal is to measure the impact of using scalable parent set selection methods and our heuristics to learn the structure of *large* Bayesian network multi-label classifiers. This is left as future work.

Appendix A

Configuration Parameters For Each Local Search Method

A.1 Greedy search

- perturbSolution: Whether to perform a perturbation of incumbent solution (Default: true)
- numPerturbationSwaps: Number of non-adjacent swaps (Default: 3)

A.2 Simulated annealing

- numRepeats: Number of repetitions at the same temperature (Default: 20, empirically determined)
- tempMax: Maximum temperature (Default: 100.0, empirically determined)
- tempMin: Minimum temperature (Default: 1.0, empirically determined)
- unchangedIterations: Maximum number of iterations without improvement until stopping the search (Default: $K/2$, empirically determined)

A.3 Tabu search

- lengthTabuList: Maximum number of unavailable movements to keep in memory (Default: $0.2n$, empirically determined)
- useAspirationCriterion: Whether to visit banned movements if that improves the incumbent solution (Default: false)

A.4 Beam search

- queueLength: Number of incumbent solutions (i.e., neighborhoods) to get at each iteration (Default: $log(n), empiricallydetermined$

A.5 ASOBS

This algorithm have the same configuration of Greedy search.

A.6 Swap search

This algorithm have the same configuration of Greedy search.

Appendix B

Empirical Results: Greedy Search

Dataset	Approach	Best Score	Avg. Best Score	Avg. Initial Score	Avg. It.
Nltcs	RND	0.3464	0.3450 ± 0.0006	0.3432 ± 0.0008	18.5200 ± 69.4781
	FAS	0.3464	0.3454 ± 0.0004	0.3436 ± 0.0000	11.9300 ± 14.0470
	DFS	0.3461	**0.3458 ± 0.0002**	**0.3443 ± 0.0002**	**10.2700 ± 2.2601**
	BFT	**0.3464**	0.3450 ± 0.0007	0.3432 ± 0.0008	11.5900 ± 16.8031
Msnbc	RND	0.0827	0.0816 ± 0.0008	0.0790 ± 0.0010	19.7500 ± 28.9550
	FAS	0.0824	0.0819 ± 0.0001	**0.0817 ± 0.0000**	**4.7800 ± 2.3337**
	DFS	0.0826	0.0814 ± 0.0007	0.0788 ± 0.0003	21.4200 ± 16.2774
	BFT	**0.0828**	**0.0820 ± 0.0006**	0.0803 ± 0.0012	29.0400 ± 72.7744
Kdd	RND	0.1447	0.1421 ± 0.0009	0.1404 ± 0.0009	**15.9100 ± 12.7684**
	FAS	**0.1463**	**0.1453 ± 0.0005**	**0.1440 ± 0.0001**	125.7000 ± 186.2100
	DFS	0.1446	0.1439 ± 0.0003	0.1428 ± 0.0004	23.0900 ± 4.2190
	BFT	0.1458	0.1434 ± 0.0009	0.1423 ± 0.0008	29.9800 ± 73.0167
Plants	RND	0.5755	0.5732 ± 0.0009	0.5711 ± 0.0010	13.4100 ± 6.9603
	FAS	0.5764	**0.5750 ± 0.0006**	0.5727 ± 0.0003	21.2600 ± 5.4784
	DFS	0.5760	0.5749 ± 0.0006	0.5711 ± 0.0006	29.6200 ± 6.1542
	BFT	**0.5767**	0.5748 ± 0.0008	**0.5739 ± 0.0008**	**8.9500 ± 3.5430**
Baudio	RND	0.1578	0.1558 ± 0.0011	0.1535 ± 0.0009	16.3400 ± 6.8094
	FAS	**0.1608**	**0.1600 ± 0.0003**	**0.1588 ± 0.0002**	25.5900 ± 5.1465
	DFS	0.1595	0.1584 ± 0.0004	0.1565 ± 0.0004	32.8500 ± 5.7020
	BFT	0.1595	0.1579 ± 0.0007	0.1569 ± 0.0007	**11.6900 ± 5.8993**
Bnetflix	RND	0.1053	0.1031 ± 0.0010	0.1014 ± 0.0009	16.3700 ± 7.1204
	FAS	**0.1082**	**0.1074 ± 0.0003**	**0.1057 ± 0.0002**	31.5900 ± 4.7992
	DFS	0.1067	0.1057 ± 0.0005	0.1042 ± 0.0005	29.4300 ± 6.3854
	BFT	0.1049	0.1025 ± 0.0010	0.1007 ± 0.0007	**15.5700 ± 8.7758**
Jester	RND	0.1440	0.1421 ± 0.0009	0.1402 ± 0.0009	**17.6800 ± 7.4140**
	FAS	**0.1467**	**0.1457 ± 0.0005**	**0.1445 ± 0.0004**	27.9300 ± 3.9421
	DFS	0.1458	0.1450 ± 0.0004	0.1439 ± 0.0005	29.8700 ± 4.1528
	BFT	0.1445	0.1420 ± 0.0009	0.1401 ± 0.0008	19.2500 ± 7.5657
Accidents	RND	0.3687	0.3593 ± 0.0042	0.3527 ± 0.0041	**14.1000 ± 6.9318**
	FAS	0.3741	0.3713 ± 0.0013	**0.3669 ± 0.0007**	26.5600 ± 4.4729
	DFS	**0.3754**	**0.3734 ± 0.0008**	0.3665 ± 0.0009	34.7300 ± 4.5966
	BFT	0.3691	0.3615 ± 0.0037	0.3560 ± 0.0033	14.6200 ± 6.1328
Tretail	RND	0.0437	0.0411 ± 0.0020	0.0406 ± 0.0021	3.8900 ± 2.5894
	FAS	0.0444	**0.0441 ± 0.0001**	**0.0438 ± 0.0000**	15.5200 ± 3.2458
	DFS	0.0442	0.0432 ± 0.0005	0.0406 ± 0.0001	28.4300 ± 4.9813
	BFT	**0.0445**	0.0436 ± 0.0008	0.0435 ± 0.0008	**3.5500 ± 2.5756**
Pumsb_star	RND	0.6880	0.6836 ± 0.0020	0.6809 ± 0.0022	**10.8000 ± 4.2020**
	FAS	0.6914	0.6893 ± 0.0010	0.6864 ± 0.0009	12.6200 ± 1.2695
	DFS	**0.6931**	**0.6924 ± 0.0004**	**0.6903 ± 0.0005**	13.2700 ± 1.1269
	BFT	0.6862	0.6811 ± 0.0029	0.6726 ± 0.0028	13.9600 ± 4.9847
Dna	RND	0.1977	0.1962 ± 0.0006	0.1958 ± 0.0006	**6.8900 ± 4.5236**
	FAS	0.1993	0.1985 ± 0.0003	0.1960 ± 0.0002	46.4700 ± 19.5466
	DFS	**0.2006**	**0.2002 ± 0.0001**	**0.1990 ± 0.0003**	53.7800 ± 47.7757
	BFT	0.1990	0.1974 ± 0.0005	0.1970 ± 0.0004	15.7800 ± 27.9612
Kosarek	RND	0.1755	0.1702 ± 0.0031	0.1659 ± 0.0031	19.5100 ± 7.6653
	FAS	0.1826	**0.1813 ± 0.0006**	**0.1792 ± 0.0006**	27.5100 ± 3.9093
	DFS	0.1815	0.1792 ± 0.0012	0.1766 ± 0.0015	40.9900 ± 4.7450
	BFT	**0.1828**	0.1791 ± 0.0022	0.1784 ± 0.0024	**10.7300 ± 6.7313**
Msweb	RND	0.1172	0.1119 ± 0.0026	0.1112 ± 0.0026	**4.1900 ± 2.4150**
	FAS	**0.1234**	**0.1222 ± 0.0005**	**0.1204 ± 0.0001**	22.8700 ± 3.6945
	DFS	0.1193	0.1154 ± 0.0021	0.1109 ± 0.0010	28.6500 ± 5.3151
	BFT	0.1231	0.1181 ± 0.0021	0.1174 ± 0.0023	5.4200 ± 2.8076
Book	RND	0.1161	0.1146 ± 0.0008	0.1143 ± 0.0008	4.0000 ± 0.1421
	FAS	**0.1222**	**0.1214 ± 0.0003**	**0.1213 ± 0.0003**	3.2900 ± 0.4777
	DFS	0.1200	0.1187 ± 0.0006	0.1185 ± 0.0006	3.3600 ± 0.5226
	BFT	0.1215	0.1203 ± 0.0005	0.1202 ± 0.0005	**2.6300 ± 0.5972**
Tmovie	RND	0.2879	0.2853 ± 0.0014	0.2849 ± 0.0014	2.8900 ± 0.6801
	FAS	**0.3037**	**0.3021 ± 0.0007**	**0.3018 ± 0.0007**	2.6000 ± 0.6667
	DFS	0.2966	0.2906 ± 0.0024	0.2899 ± 0.0025	3.4100 ± 0.5522
	BFT	0.2985	0.2954 ± 0.0011	0.2951 ± 0.0010	**2.4200 ± 0.5160**
Cwebkb	RND	0.1122	0.1093 ± 0.0013	0.1092 ± 0.0013	2.7300 ± 0.4683
	FAS	**0.1235**	**0.1222 ± 0.0005**	**0.1219 ± 0.0005**	2.8000 ± 0.4264
	DFS	0.1216	0.1198 ± 0.0008	0.1196 ± 0.0008	2.6100 ± 0.5104
	BFT	0.1196	0.1177 ± 0.0011	0.1176 ± 0.0011	**2.1500 ± 0.4352**
Cr52	RND	0.1443	0.1345 ± 0.0042	0.1341 ± 0.0042	2.4700 ± 0.5214
	FAS	**0.1595**	**0.1574 ± 0.0011**	**0.1571 ± 0.0011**	2.2200 ± 0.4163
	DFS	0.1573	0.1559 ± 0.0007	0.1555 ± 0.0007	2.0800 ± 0.2727
	BFT	0.1547	0.1488 ± 0.0029	0.1486 ± 0.0029	**2.0000 ± 0.1421**
C20ng	RND	0.0703	0.0682 ± 0.0010	0.0682 ± 0.0010	1.0100 ± 0.1000
	FAS	**0.0769**	**0.0763 ± 0.0003**	**0.0763 ± 0.0003**	1.0100 ± 0.1000
	DFS	0.0753	0.0742 ± 0.0005	0.0741 ± 0.0005	1.0100 ± 0.1000
	BFT	0.0747	0.0735 ± 0.0007	0.0735 ± 0.0007	**1.0100 ± 0.1000**
Bbc	RND	0.0680	0.0652 ± 0.0010	0.0649 ± 0.0009	6.7400 ± 0.8718
	FAS	**0.0772**	**0.0764 ± 0.0003**	**0.0761 ± 0.0003**	**6.0200 ± 0.6510**
	DFS	0.0733	0.0718 ± 0.0007	0.0714 ± 0.0007	7.0700 ± 0.3828
	BFT	0.0732	0.0719 ± 0.0006	0.0718 ± 0.0005	6.6100 ± 0.9837
Ad	RND	0.6917	0.6850 ± 0.0024	0.6846 ± 0.0023	**2.1100 ± 1.4347**
	FAS	0.7005	0.6961 ± 0.0020	0.6951 ± 0.0019	5.1900 ± 0.5449
	DFS	**0.7032**	**0.6972 ± 0.0022**	**0.6960 ± 0.0022**	5.4800 ± 0.5942
	BFT	0.6967	0.6915 ± 0.0026	0.6911 ± 0.0025	3.3500 ± 1.7078

Table B.1: *Best score obtained, Average best score obtained, Average initial score generated, Average number of iterations (Avg. It.) using sequential selection (best values in bold)*

Dataset	Approach	Best Score	Avg. Best Score	Avg. Initial Score	Avg. It.
Nltcs	RND	0.3416	0.3378 ± 0.0020	0.3327 ± 0.0029	6.3700 ± 3.1676
	FAS	0.3421	**0.3412 ± 0.0005**	**0.3378 ± 0.0010**	**4.8400 ± 1.8462**
	DFS	**0.3428**	0.3409 ± 0.0006	0.3373 ± 0.0008	5.3400 ± 1.4086
	BFT	0.3417	0.3377 ± 0.0022	0.3322 ± 0.0028	6.5400 ± 2.9997
Msnbc	RND	**0.0592**	0.0553 ± 0.0020	0.0501 ± 0.0030	21.6000 ± 69.9958
	FAS	0.0585	**0.0567 ± 0.0005**	0.0540 ± 0.0000	5.8900 ± 5.4102
	DFS	0.0586	0.0554 ± 0.0010	**0.0543 ± 0.0005**	**3.5500 ± 2.3241**
	BFT	0.0588	0.0565 ± 0.0014	0.0519 ± 0.0031	42.1400 ± 109.5346
Kdd	RND	0.1399	0.1376 ± 0.0010	0.1357 ± 0.0012	**10.3800 ± 5.5628**
	FAS	**0.1417**	**0.1402 ± 0.0006**	**0.1388 ± 0.0002**	12.5800 ± 3.9059
	DFS	0.1409	0.1394 ± 0.0008	0.1371 ± 0.0002	24.0000 ± 5.0772
	BFT	0.1417	0.1391 ± 0.0012	0.1373 ± 0.0012	11.7600 ± 10.3652
Plants	RND	0.5646	0.5607 ± 0.0019	0.5564 ± 0.0020	11.6500 ± 4.7234
	FAS	**0.5672**	**0.5645 ± 0.0009**	**0.5616 ± 0.0006**	19.2100 ± 50.4085
	DFS	0.5667	0.5639 ± 0.0010	0.5574 ± 0.0009	26.2300 ± 5.5738
	BFT	0.5660	0.5622 ± 0.0016	0.5587 ± 0.0017	**11.5100 ± 4.7450**
Baudio	RND	0.1575	0.1558 ± 0.0009	0.1537 ± 0.0009	15.4200 ± 6.7453
	FAS	0.1605	0.1590 ± 0.0006	0.1567 ± 0.0006	33.3200 ± 34.8109
	DFS	**0.1609**	**0.1599 ± 0.0004**	**0.1575 ± 0.0003**	31.7400 ± 4.6092
	BFT	0.1600	0.1580 ± 0.0007	0.1567 ± 0.0007	**12.2000 ± 5.3333**
Bnetflix	RND	0.1044	0.1024 ± 0.0008	0.1007 ± 0.0007	14.9400 ± 5.6726
	FAS	0.1056	0.1042 ± 0.0005	0.1022 ± 0.0004	24.0000 ± 5.5741
	DFS	**0.1063**	**0.1055 ± 0.0004**	**0.1038 ± 0.0005**	30.9100 ± 6.0354
	BFT	0.1040	0.1018 ± 0.0009	0.1003 ± 0.0006	**14.1700 ± 7.5653**
Jester	RND	0.1438	0.1416 ± 0.0010	0.1392 ± 0.0010	18.2300 ± 9.1142
	FAS	**0.1468**	**0.1459 ± 0.0003**	**0.1446 ± 0.0003**	21.5600 ± 4.0059
	DFS	0.1463	0.1454 ± 0.0004	0.1434 ± 0.0005	41.3200 ± 6.4242
	BFT	0.1447	0.1418 ± 0.0011	0.1398 ± 0.0007	**17.3100 ± 7.5755**
Accidents	RND	0.3787	0.3700 ± 0.0041	0.3641 ± 0.0041	**12.1600 ± 6.0431**
	FAS	0.3847	0.3794 ± 0.0024	0.3710 ± 0.0016	26.1100 ± 5.7417
	DFS	**0.3853**	**0.3826 ± 0.0012**	**0.3734 ± 0.0006**	37.9200 ± 6.7952
	BFT	0.3802	0.3713 ± 0.0034	0.3646 ± 0.0033	14.0600 ± 5.6261
Tretail	RND	0.0435	0.0422 ± 0.0007	0.0419 ± 0.0008	3.9100 ± 2.6173
	FAS	0.0442	**0.0440 ± 0.0001**	0.0430 ± 0.0001	15.9900 ± 2.7943
	DFS	0.0441	0.0436 ± 0.0002	0.0424 ± 0.0002	25.9500 ± 4.2814
	BFT	**0.0444**	0.0436 ± 0.0005	**0.0435 ± 0.0005**	**3.6000 ± 2.5859**
Pumsb_star	RND	0.6688	0.6587 ± 0.0050	0.6514 ± 0.0059	**11.3400 ± 5.0075**
	FAS	0.6747	0.6702 ± 0.0020	0.6661 ± 0.0019	21.6800 ± 4.5213
	DFS	**0.6786**	**0.6750 ± 0.0014**	**0.6700 ± 0.0021**	24.6900 ± 3.9739
	BFT	0.6739	0.6595 ± 0.0057	0.6502 ± 0.0044	12.4300 ± 6.5771
Dna	RND	0.1990	0.1950 ± 0.0011	0.1941 ± 0.0007	**15.0400 ± 53.3405**
	FAS	0.1992	0.1980 ± 0.0003	0.1954 ± 0.0002	47.5500 ± 46.2949
	DFS	**0.2009**	**0.2006 ± 0.0001**	**0.1990 ± 0.0002**	60.1200 ± 44.8261
	BFT	0.1995	0.1968 ± 0.0009	0.1962 ± 0.0005	23.4400 ± 62.3263
Kosarek	RND	0.1778	0.1710 ± 0.0031	0.1670 ± 0.0033	20.0500 ± 12.1642
	FAS	0.1807	0.1759 ± 0.0014	0.1707 ± 0.0005	49.6800 ± 27.6025
	DFS	0.1796	0.1756 ± 0.0019	0.1667 ± 0.0007	64.8500 ± 16.5946
	BFT	**0.1816**	**0.1786 ± 0.0017**	**0.1776 ± 0.0020**	**14.7500 ± 14.4567**
Msweb	RND	0.1177	0.1120 ± 0.0023	0.1113 ± 0.0022	**3.4200 ± 3.0983**
	FAS	**0.1221**	**0.1196 ± 0.0008**	**0.1170 ± 0.0001**	26.6400 ± 4.7748
	DFS	0.1199	0.1150 ± 0.0020	0.1097 ± 0.0007	34.6200 ± 5.4989
	BFT	0.1209	0.1171 ± 0.0019	0.1165 ± 0.0018	4.9100 ± 2.8144
Book	RND	0.1214	0.1195 ± 0.0008	0.1188 ± 0.0008	11.4800 ± 3.7564
	FAS	**0.1269**	**0.1263 ± 0.0003**	**0.1257 ± 0.0002**	16.5000 ± 3.1988
	DFS	0.1210	0.1195 ± 0.0006	0.1184 ± 0.0004	17.0400 ± 3.2750
	BFT	0.1262	0.1250 ± 0.0005	0.1249 ± 0.0005	**8.0100 ± 2.5564**
Tmovie	RND	0.3169	0.3123 ± 0.0016	0.3113 ± 0.0016	7.6600 ± 1.5777
	FAS	**0.3275**	**0.3255 ± 0.0010**	**0.3245 ± 0.0010**	6.5300 ± 1.1142
	DFS	0.3135	0.3093 ± 0.0016	0.3067 ± 0.0017	11.5500 ± 1.9611
	BFT	0.3259	0.3235 ± 0.0013	0.3230 ± 0.0013	**5.7800 ± 0.9490**
Cwebkb	RND	0.1268	0.1245 ± 0.0010	0.1241 ± 0.0010	6.6200 ± 0.6321
	FAS	**0.1341**	**0.1335 ± 0.0003**	**0.1331 ± 0.0003**	6.6000 ± 0.7247
	DFS	0.1229	0.1210 ± 0.0010	0.1201 ± 0.0010	7.4500 ± 0.8689
	BFT	0.1337	0.1324 ± 0.0007	0.1322 ± 0.0007	**5.3400 ± 0.9235**
Cr52	RND	0.1682	0.1611 ± 0.0035	0.1606 ± 0.0034	4.8900 ± 0.6497
	FAS	**0.1799**	**0.1787 ± 0.0004**	**0.1781 ± 0.0004**	4.8500 ± 0.3589
	DFS	0.1616	0.1500 ± 0.0028	0.1475 ± 0.0010	6.7000 ± 0.5025
	BFT	0.1787	0.1751 ± 0.0019	0.1748 ± 0.0019	**4.0300 ± 0.1714**
C20ng	RND	0.0890	0.0859 ± 0.0011	0.0857 ± 0.0011	2.0100 ± 0.1000
	FAS	0.0944	**0.0939 ± 0.0002**	**0.0938 ± 0.0002**	2.0100 ± 0.1000
	DFS	0.0856	0.0844 ± 0.0005	0.0842 ± 0.0005	2.0300 ± 0.2227
	BFT	**0.0951**	0.0937 ± 0.0007	0.0937 ± 0.0006	**2.0100 ± 0.1000**
Bbc	RND	0.0754	0.0735 ± 0.0008	0.0733 ± 0.0007	6.8100 ± 0.8127
	FAS	**0.0805**	**0.0802 ± 0.0002**	**0.0799 ± 0.0002**	6.7500 ± 0.6093
	DFS	0.0751	0.0737 ± 0.0006	0.0733 ± 0.0006	7.8900 ± 0.6340
	BFT	0.0771	0.0761 ± 0.0005	0.0759 ± 0.0005	**6.0700 ± 0.7555**
Ad	RND	0.7121	0.7069 ± 0.0019	0.7066 ± 0.0018	**2.3300 ± 1.4843**
	FAS	0.7149	0.7115 ± 0.0016	0.7104 ± 0.0015	5.8000 ± 0.5685
	DFS	**0.7177**	**0.7144 ± 0.0015**	**0.7134 ± 0.0015**	5.8700 ± 0.5056
	BFT	0.7132	0.7100 ± 0.0016	0.7097 ± 0.0016	3.4700 ± 1.6046

Table B.2: *Best score obtained, Average best score obtained, Average initial score generated, Average number of iterations (Avg. It.) using greedy selection (best values in bold)*

Dataset	Approach	Best Score	Avg. Best Score	Avg. Initial Score	Avg. It.
Nltcs	RND	0.3462	0.3449 ± 0.0006	0.3433 ± 0.0009	**7.7000 ± 3.7457**
	FAS	0.3460	0.3453 ± 0.0003	0.3436 ± 0.0000	9.9500 ± 3.1826
	DFS	**0.3463**	**0.3459 ± 0.0002**	**0.3443 ± 0.0002**	14.5000 ± 35.4730
	BFT	0.3461	0.3450 ± 0.0006	0.3432 ± 0.0008	7.9300 ± 4.7103
Msnbc	RND	0.0829	**0.0809 ± 0.0009**	0.0749 ± 0.0038	18.1800 ± 17.4729
	FAS	0.0822	0.0804 ± 0.0007	**0.0771 ± 0.0000**	16.4800 ± 31.0417
	DFS	0.0830	0.0808 ± 0.0008	0.0760 ± 0.0014	18.4900 ± 8.5770
	BFT	**0.0832**	0.0808 ± 0.0011	0.0741 ± 0.0036	**15.0500 ± 8.6403**
Kdd	RND	0.1363	0.1299 ± 0.0026	**0.1261 ± 0.0029**	8.5700 ± 4.6368
	FAS	0.1357	**0.1326 ± 0.0016**	0.1231 ± 0.0005	14.7500 ± 4.1497
	DFS	**0.1373**	0.1323 ± 0.0022	0.1181 ± 0.0002	23.2100 ± 6.0491
	BFT	0.1351	0.1302 ± 0.0026	0.1258 ± 0.0031	**7.3300 ± 3.5278**
Plants	RND	0.5438	0.5330 ± 0.0048	0.5226 ± 0.0050	9.9000 ± 4.7280
	FAS	0.5436	0.5359 ± 0.0035	0.5208 ± 0.0018	19.1000 ± 5.5167
	DFS	**0.5507**	**0.5468 ± 0.0012**	**0.5350 ± 0.0018**	34.7300 ± 5.1598
	BFT	0.5422	0.5331 ± 0.0039	0.5259 ± 0.0042	**9.7200 ± 5.8948**
Baudio	RND	0.1409	0.1338 ± 0.0028	0.1291 ± 0.0027	9.4300 ± 6.8744
	FAS	0.1448	**0.1402 ± 0.0016**	0.1310 ± 0.0003	23.8700 ± 4.6159
	DFS	0.1419	0.1370 ± 0.0022	0.1219 ± 0.0013	26.8300 ± 5.9341
	BFT	**0.1460**	0.1392 ± 0.0025	**0.1375 ± 0.0025**	**7.1500 ± 5.4279**
Bnetflix	RND	0.0942	0.0896 ± 0.0021	0.0867 ± 0.0020	8.2100 ± 3.8067
	FAS	**0.0968**	**0.0940 ± 0.0011**	0.0872 ± 0.0002	22.4700 ± 4.9061
	DFS	0.0953	0.0920 ± 0.0014	0.0807 ± 0.0013	26.5400 ± 5.7162
	BFT	0.0963	0.0912 ± 0.0018	**0.0894 ± 0.0016**	**7.4700 ± 4.7831**
Jester	RND	0.1272	0.1197 ± 0.0036	0.1143 ± 0.0032	**13.5000 ± 11.1305**
	FAS	0.1328	**0.1272 ± 0.0018**	**0.1201 ± 0.0004**	23.9700 ± 22.8824
	DFS	**0.1347**	0.1235 ± 0.0036	0.1076 ± 0.0014	46.4400 ± 81.2328
	BFT	0.1287	0.1227 ± 0.0031	0.1198 ± 0.0027	13.8400 ± 12.4143
Accidents	RND	0.3233	0.3086 ± 0.0075	0.2995 ± 0.0078	8.9700 ± 6.3936
	FAS	**0.3366**	**0.3298 ± 0.0031**	**0.3204 ± 0.0007**	17.5800 ± 4.7080
	DFS	0.3339	0.3269 ± 0.0036	0.3044 ± 0.0012	27.9100 ± 8.9782
	BFT	0.3240	0.3111 ± 0.0064	0.3049 ± 0.0064	**7.2800 ± 4.7013**
Tretail	RND	0.0435	0.0408 ± 0.0019	0.0400 ± 0.0020	4.3500 ± 2.6756
	FAS	0.0443	**0.0440 ± 0.0001**	**0.0436 ± 0.0000**	15.3500 ± 3.0231
	DFS	0.0439	0.0429 ± 0.0005	0.0365 ± 0.0001	25.3400 ± 4.9934
	BFT	**0.0444**	0.0435 ± 0.0005	0.0434 ± 0.0006	**3.5700 ± 2.2841**
Pumsb_star	RND	0.6296	0.6114 ± 0.0066	0.6042 ± 0.0071	**8.6400 ± 5.2887**
	FAS	0.6424	0.6365 ± 0.0024	0.6301 ± 0.0015	11.2700 ± 3.1586
	DFS	**0.6431**	**0.6396 ± 0.0018**	**0.6337 ± 0.0022**	19.6900 ± 9.5755
	BFT	0.6238	0.6093 ± 0.0062	0.6000 ± 0.0045	9.5700 ± 5.1488
Dna	RND	0.1942	0.1891 ± 0.0013	0.1883 ± 0.0006	**19.2900 ± 80.7738**
	FAS	**0.1963**	**0.1960 ± 0.0001**	**0.1945 ± 0.0001**	52.1400 ± 14.4614
	DFS	0.1962	0.1957 ± 0.0003	0.1924 ± 0.0008	32.7000 ± 5.1864
	BFT	0.1950	0.1915 ± 0.0012	0.1905 ± 0.0006	38.3300 ± 83.4179
Kosarek	RND	0.1739	0.1628 ± 0.0050	0.1577 ± 0.0047	17.6100 ± 10.1473
	FAS	0.1770	**0.1745 ± 0.0011**	0.1698 ± 0.0009	43.0000 ± 17.3927
	DFS	0.1747	0.1675 ± 0.0035	0.1535 ± 0.0015	61.5000 ± 20.8445
	BFT	**0.1794**	0.1729 ± 0.0028	**0.1716 ± 0.0028**	**15.7600 ± 44.1746**
Msweb	RND	0.1143	0.1069 ± 0.0032	0.1057 ± 0.0031	**4.0000 ± 4.9604**
	FAS	**0.1204**	**0.1189 ± 0.0006**	**0.1166 ± 0.0000**	21.4000 ± 3.5505
	DFS	0.1159	0.1099 ± 0.0028	0.0943 ± 0.0025	33.6700 ± 5.1797
	BFT	0.1199	0.1141 ± 0.0029	0.1130 ± 0.0029	4.5200 ± 2.5285
Book	RND	0.1176	0.1152 ± 0.0012	0.1145 ± 0.0011	9.0700 ± 3.8619
	FAS	**0.1241**	**0.1233 ± 0.0003**	0.1211 ± 0.0003	54.5400 ± 6.9956
	DFS	0.1202	0.1165 ± 0.0010	0.1108 ± 0.0006	59.4300 ± 11.4550
	BFT	0.1238	0.1220 ± 0.0007	**0.1217 ± 0.0007**	**6.5800 ± 3.1178**
Tmovie	RND	0.2913	0.2810 ± 0.0048	0.2766 ± 0.0042	17.7100 ± 12.9555
	FAS	**0.3065**	**0.3001 ± 0.0023**	0.2876 ± 0.0005	56.6500 ± 10.2330
	DFS	0.2929	0.2823 ± 0.0042	0.2516 ± 0.0040	70.6100 ± 10.9313
	BFT	0.3064	0.2993 ± 0.0030	**0.2974 ± 0.0031**	**13.7500 ± 9.4798**
Cwebkb	RND	0.1210	0.1172 ± 0.0015	0.1165 ± 0.0015	9.6800 ± 3.9180
	FAS	0.1281	**0.1270 ± 0.0004**	0.1246 ± 0.0001	18.6700 ± 2.4332
	DFS	0.1173	0.1141 ± 0.0012	0.1103 ± 0.0008	19.7700 ± 1.9533
	BFT	**0.1289**	0.1265 ± 0.0013	**0.1262 ± 0.0013**	**6.1400 ± 3.4142**
Cr52	RND	0.1560	0.1462 ± 0.0048	0.1447 ± 0.0049	10.7000 ± 4.0088
	FAS	0.1679	**0.1656 ± 0.0007**	**0.1629 ± 0.0003**	18.8800 ± 0.7004
	DFS	0.1505	0.1369 ± 0.0046	0.1309 ± 0.0030	18.4600 ± 0.7839
	BFT	**0.1685**	0.1628 ± 0.0034	0.1622 ± 0.0034	**6.1200 ± 2.4793**
C20ng	RND	0.0765	0.0725 ± 0.0021	0.0718 ± 0.0020	8.9400 ± 2.4155
	FAS	0.0838	0.0833 ± 0.0002	0.0822 ± 0.0001	11.0800 ± 0.5257
	DFS	0.0657	0.0617 ± 0.0013	0.0595 ± 0.0007	11.1800 ± 1.0188
	BFT	**0.0870**	**0.0841 ± 0.0016**	**0.0839 ± 0.0016**	**6.1800 ± 2.4919**
Bbc	RND	0.0730	0.0711 ± 0.0009	0.0707 ± 0.0009	10.9000 ± 2.7136
	FAS	**0.0800**	**0.0797 ± 0.0002**	**0.0789 ± 0.0001**	12.7900 ± 0.9566
	DFS	0.0749	0.0723 ± 0.0011	0.0714 ± 0.0011	13.1000 ± 0.7035
	BFT	0.0759	0.0742 ± 0.0006	0.0739 ± 0.0006	**10.1800 ± 3.4594**
Ad	RND	0.7017	0.6936 ± 0.0026	0.6932 ± 0.0024	**2.0400 ± 1.4834**
	FAS	0.7069	**0.7036 ± 0.0015**	**0.7013 ± 0.0014**	5.9800 ± 0.5121
	DFS	**0.7073**	0.7025 ± 0.0020	0.7003 ± 0.0020	5.7000 ± 0.5222
	BFT	0.7043	0.6995 ± 0.0019	0.6992 ± 0.0019	3.2500 ± 1.7019

Table B.3: *Best score obtained, Average best score obtained, Average initial score generated, Average number of iterations (Avg. It.) using independence selection (best values in bold)*

Appendix C

Empirical Results: Simulated Annealing

Dataset	Approach	Best Score	Avg. Best Score	Avg. Initial Score	Avg. It.
Nltcs	RND	**0.3466**	0.3466 ± 0.0001	0.3431 ± 0.0009	492.5600 ± 19.0588
	FAS	0.3466	**0.3466 ± 0.0001**	0.3436 ± 0.0000	494.7800 ± 18.9937
	DFS	0.3466	0.3466 ± 0.0001	**0.3442 ± 0.0002**	**489.0700 ± 37.6419**
	BFT	0.3466	0.3466 ± 0.0001	0.3432 ± 0.0008	493.4700 ± 22.8701
Msnbc	RND	**0.0828**	0.0825 ± 0.0003	0.0788 ± 0.0010	470.8500 ± 55.6623
	FAS	0.0827	0.0824 ± 0.0001	**0.0817 ± 0.0000**	**470.4400 ± 59.5967**
	DFS	0.0827	0.0824 ± 0.0002	0.0789 ± 0.0003	485.3900 ± 40.0204
	BFT	0.0828	**0.0825 ± 0.0002**	0.0804 ± 0.0011	473.8300 ± 52.7556
Kdd	RND	0.1455	0.1440 ± 0.0008	0.1404 ± 0.0009	500.0000 ± 0.0000
	FAS	**0.1470**	**0.1462 ± 0.0004**	**0.1440 ± 0.0001**	**498.3600 ± 16.4000**
	DFS	0.1463	0.1456 ± 0.0004	0.1428 ± 0.0004	500.0000 ± 0.0000
	BFT	0.1467	0.1455 ± 0.0006	0.1423 ± 0.0008	500.0000 ± 0.0000
Plants	RND	0.5757	0.5739 ± 0.0009	0.5710 ± 0.0010	195.2400 ± 19.8678
	FAS	0.5764	0.5752 ± 0.0005	0.5726 ± 0.0003	**181.5100 ± 21.7393**
	DFS	0.5772	0.5760 ± 0.0006	0.5710 ± 0.0006	190.6100 ± 17.8217
	BFT	**0.5776**	**0.5761 ± 0.0006**	**0.5739 ± 0.0007**	187.1300 ± 15.2545
Baudio	RND	0.1587	0.1554 ± 0.0011	0.1535 ± 0.0010	186.7600 ± 23.9693
	FAS	**0.1615**	**0.1606 ± 0.0005**	**0.1588 ± 0.0002**	189.5400 ± 24.3462
	DFS	0.1604	0.1589 ± 0.0005	0.1564 ± 0.0004	184.0600 ± 22.8384
	BFT	0.1602	0.1584 ± 0.0008	0.1568 ± 0.0008	**173.4700 ± 14.8484**
Bnetflix	RND	0.1062	0.1039 ± 0.0009	0.1013 ± 0.0008	**369.7000 ± 41.5970**
	FAS	**0.1090**	**0.1082 ± 0.0003**	**0.1057 ± 0.0002**	385.2200 ± 38.8464
	DFS	0.1084	0.1076 ± 0.0004	0.1041 ± 0.0006	433.4600 ± 43.1723
	BFT	0.1057	0.1035 ± 0.0008	0.1007 ± 0.0008	416.4700 ± 44.3736
Jester	RND	0.1432	0.1410 ± 0.0009	0.1402 ± 0.0009	96.8700 ± 12.5470
	FAS	**0.1464**	**0.1451 ± 0.0005**	**0.1444 ± 0.0004**	**90.8800 ± 9.9375**
	DFS	0.1456	0.1445 ± 0.0005	0.1439 ± 0.0006	104.9600 ± 8.6421
	BFT	0.1430	0.1411 ± 0.0007	0.1400 ± 0.0007	101.4900 ± 10.9595
Accidents	RND	0.3714	0.3627 ± 0.0041	0.3532 ± 0.0046	500.0000 ± 0.0000
	FAS	**0.3776**	0.3748 ± 0.0012	**0.3670 ± 0.0008**	500.0000 ± 0.0000
	DFS	0.3770	**0.3757 ± 0.0007**	0.3665 ± 0.0010	500.0000 ± 0.0000
	BFT	0.3714	0.3648 ± 0.0034	0.3562 ± 0.0037	**500.0000 ± 0.0000**
Tretail	RND	0.0437	0.0413 ± 0.0019	0.0405 ± 0.0020	475.3500 ± 68.6120
	FAS	0.0443	**0.0439 ± 0.0001**	**0.0438 ± 0.0000**	**279.5800 ± 72.8089**
	DFS	0.0435	0.0428 ± 0.0003	0.0406 ± 0.0001	484.4400 ± 57.2025
	BFT	**0.0447**	0.0438 ± 0.0008	0.0435 ± 0.0009	436.7800 ± 100.1384
Pumsb_star	RND	0.6869	0.6820 ± 0.0021	0.6807 ± 0.0023	72.1200 ± 9.1710
	FAS	0.6909	0.6890 ± 0.0011	0.6865 ± 0.0009	67.7700 ± 9.2582
	DFS	**0.6927**	**0.6920 ± 0.0004**	**0.6903 ± 0.0005**	**55.4500 ± 5.2114**
	BFT	0.6810	0.6749 ± 0.0027	0.6731 ± 0.0026	77.5400 ± 10.3019
Dna	RND	0.1977	0.1963 ± 0.0006	0.1958 ± 0.0006	470.6200 ± 70.2676
	FAS	0.1993	0.1984 ± 0.0005	0.1961 ± 0.0002	493.9500 ± 34.7735
	DFS	**0.1996**	**0.1990 ± 0.0003**	**0.1990 ± 0.0003**	**251.0100 ± 0.1000**
	BFT	0.1981	0.1973 ± 0.0004	0.1971 ± 0.0004	433.6400 ± 93.3050
Kosarek	RND	0.1746	0.1672 ± 0.0037	0.1658 ± 0.0036	157.0200 ± 20.5210
	FAS	0.1814	**0.1804 ± 0.0006**	**0.1791 ± 0.0007**	155.9800 ± 18.5080
	DFS	0.1805	0.1780 ± 0.0012	0.1765 ± 0.0014	204.5000 ± 18.6523
	BFT	**0.1824**	0.1788 ± 0.0025	0.1783 ± 0.0025	**133.7500 ± 19.1825**
Msweb	RND	0.1170	0.1122 ± 0.0022	0.1113 ± 0.0022	487.7500 ± 46.1030
	FAS	0.1233	**0.1219 ± 0.0006**	**0.1204 ± 0.0001**	474.8500 ± 72.3838
	DFS	0.1162	0.1142 ± 0.0009	0.1110 ± 0.0011	499.7500 ± 2.5000
	BFT	**0.1234**	0.1184 ± 0.0025	0.1173 ± 0.0023	**472.2900 ± 73.2846**
Book	RND	0.1158	0.1145 ± 0.0007	0.1144 ± 0.0007	45.4000 ± 3.6735
	FAS	**0.1219**	**0.1213 ± 0.0003**	**0.1213 ± 0.0003**	44.1100 ± 2.9298
	DFS	0.1195	0.1184 ± 0.0006	0.1184 ± 0.0006	47.5700 ± 2.9721
	BFT	0.1214	0.1202 ± 0.0005	0.1202 ± 0.0005	**39.9900 ± 3.9835**
Tmovie	RND	0.2877	0.2850 ± 0.0012	0.2849 ± 0.0012	40.3300 ± 1.8913
	FAS	**0.3035**	**0.3018 ± 0.0007**	**0.3018 ± 0.0007**	36.8100 ± 2.4400
	DFS	0.2960	0.2899 ± 0.0029	0.2898 ± 0.0028	46.3300 ± 2.1745
	BFT	0.2974	0.2951 ± 0.0011	0.2949 ± 0.0010	**32.6200 ± 1.2373**
Cwebkb	RND	0.1117	0.1089 ± 0.0014	0.1089 ± 0.0013	53.1600 ± 1.7963
	FAS	**0.1231**	**0.1219 ± 0.0005**	**0.1219 ± 0.0005**	51.0200 ± 2.7924
	DFS	0.1210	0.1196 ± 0.0009	0.1196 ± 0.0009	50.0100 ± 1.5341
	BFT	0.1198	0.1176 ± 0.0010	0.1175 ± 0.0010	**33.2600 ± 1.6734**
Cr52	RND	0.1438	0.1348 ± 0.0044	0.1347 ± 0.0043	46.8600 ± 3.6405
	FAS	**0.1592**	**0.1570 ± 0.0011**	**0.1569 ± 0.0011**	43.2000 ± 3.3333
	DFS	0.1575	0.1556 ± 0.0009	0.1556 ± 0.0008	41.4800 ± 1.9770
	BFT	0.1537	0.1487 ± 0.0029	0.1486 ± 0.0029	**34.1500 ± 1.7545**
C20ng	RND	0.0702	0.0680 ± 0.0010	0.0680 ± 0.0010	7.3900 ± 0.6801
	FAS	**0.0769**	**0.0763 ± 0.0003**	**0.0763 ± 0.0003**	7.2700 ± 0.8270
	DFS	0.0750	0.0742 ± 0.0005	0.0741 ± 0.0004	7.2300 ± 0.6333
	BFT	0.0748	0.0733 ± 0.0006	0.0733 ± 0.0006	**6.1800 ± 0.4795**
Bbc	RND	0.0669	0.0652 ± 0.0009	0.0651 ± 0.0009	179.4300 ± 4.1469
	FAS	**0.0768**	**0.0762 ± 0.0003**	**0.0762 ± 0.0003**	**165.2600 ± 3.1738**
	DFS	0.0732	0.0715 ± 0.0007	0.0715 ± 0.0007	179.2100 ± 3.0987
	BFT	0.0730	0.0718 ± 0.0006	0.0718 ± 0.0006	170.8700 ± 2.9359
Ad	RND	0.6903	0.6849 ± 0.0021	0.6847 ± 0.0020	**205.7100 ± 9.5709**
	FAS	0.7006	0.6950 ± 0.0021	0.6949 ± 0.0020	227.0400 ± 12.4405
	DFS	**0.7014**	**0.6959 ± 0.0021**	**0.6958 ± 0.0021**	207.2100 ± 6.8848
	BFT	0.6971	0.6915 ± 0.0023	0.6913 ± 0.0023	219.8400 ± 6.5485

Table C.1: *Best score obtained, Average best score obtained, Average initial score generated, Average number of iterations (Avg. It.) using sequential selection (best values in bold)*

Dataset	Approach	Best Score	Avg. Best Score	Avg. Initial Score	Avg. It.
Nltcs	RND	**0.3439**	0.3429 ± 0.0007	0.3332 ± 0.0029	447.1000 ± 63.9339
	FAS	0.3439	**0.3431 ± 0.0006**	**0.3378 ± 0.0010**	**428.0600 ± 90.8552**
	DFS	0.3439	0.3430 ± 0.0006	0.3372 ± 0.0007	432.8400 ± 90.6698
	BFT	0.3439	0.3428 ± 0.0008	0.3326 ± 0.0029	454.1000 ± 65.1820
Msnbc	RND	0.0593	0.0562 ± 0.0020	0.0496 ± 0.0031	**434.2200 ± 83.1239**
	FAS	0.0587	0.0577 ± 0.0004	0.0540 ± 0.0000	454.4500 ± 82.0215
	DFS	0.0594	**0.0583 ± 0.0008**	**0.0544 ± 0.0006**	484.7100 ± 51.0665
	BFT	**0.0595**	0.0570 ± 0.0017	0.0520 ± 0.0029	440.9800 ± 88.9897
Kdd	RND	0.1421	0.1403 ± 0.0009	0.1356 ± 0.0012	500.0000 ± 0.0000
	FAS	**0.1439**	**0.1423 ± 0.0006**	**0.1388 ± 0.0002**	500.0000 ± 0.0000
	DFS	0.1426	0.1413 ± 0.0006	0.1371 ± 0.0002	500.0000 ± 0.0000
	BFT	0.1437	0.1417 ± 0.0009	0.1372 ± 0.0014	**500.0000 ± 0.0000**
Plants	RND	0.5692	0.5658 ± 0.0015	0.5564 ± 0.0019	500.0000 ± 0.0000
	FAS	0.5694	0.5675 ± 0.0009	**0.5615 ± 0.0006**	500.0000 ± 0.0000
	DFS	**0.5708**	**0.5685 ± 0.0010**	0.5573 ± 0.0009	500.0000 ± 0.0000
	BFT	0.5702	0.5676 ± 0.0014	0.5587 ± 0.0018	**500.0000 ± 0.0000**
Baudio	RND	0.1600	0.1581 ± 0.0009	0.1539 ± 0.0010	500.0000 ± 0.0000
	FAS	0.1615	0.1603 ± 0.0005	0.1566 ± 0.0006	500.0000 ± 0.0000
	DFS	**0.1632**	**0.1624 ± 0.0005**	**0.1575 ± 0.0003**	500.0000 ± 0.0000
	BFT	0.1616	0.1603 ± 0.0007	0.1567 ± 0.0007	**500.0000 ± 0.0000**
Bnetflix	RND	0.1058	0.1040 ± 0.0008	0.1006 ± 0.0008	500.0000 ± 0.0000
	FAS	0.1061	0.1051 ± 0.0004	0.1022 ± 0.0004	500.0000 ± 0.0000
	DFS	**0.1086**	**0.1075 ± 0.0005**	**0.1038 ± 0.0005**	500.0000 ± 0.0000
	BFT	0.1045	0.1034 ± 0.0005	0.1004 ± 0.0006	**500.0000 ± 0.0000**
Jester	RND	0.1452	0.1433 ± 0.0009	0.1394 ± 0.0011	500.0000 ± 0.0000
	FAS	**0.1481**	**0.1474 ± 0.0003**	**0.1446 ± 0.0003**	500.0000 ± 0.0000
	DFS	0.1479	0.1472 ± 0.0003	0.1434 ± 0.0005	500.0000 ± 0.0000
	BFT	0.1456	0.1442 ± 0.0006	0.1399 ± 0.0006	**500.0000 ± 0.0000**
Accidents	RND	0.3835	0.3738 ± 0.0035	0.3642 ± 0.0042	500.0000 ± 0.0000
	FAS	0.3883	0.3836 ± 0.0019	0.3708 ± 0.0013	500.0000 ± 0.0000
	DFS	**0.3893**	**0.3854 ± 0.0014**	**0.3734 ± 0.0005**	500.0000 ± 0.0000
	BFT	0.3792	0.3733 ± 0.0029	0.3641 ± 0.0033	**500.0000 ± 0.0000**
Tretail	RND	0.0436	0.0425 ± 0.0007	0.0420 ± 0.0008	460.6700 ± 83.0077
	FAS	0.0443	**0.0439 ± 0.0002**	0.0430 ± 0.0001	**377.3600 ± 113.5651**
	DFS	0.0439	0.0435 ± 0.0002	0.0424 ± 0.0002	483.0200 ± 55.9267
	BFT	**0.0445**	0.0437 ± 0.0005	**0.0435 ± 0.0005**	414.0400 ± 109.6187
Pumsb_star	RND	0.6695	0.6606 ± 0.0043	0.6512 ± 0.0050	500.0000 ± 0.0000
	FAS	0.6756	0.6706 ± 0.0016	0.6662 ± 0.0019	500.0000 ± 0.0000
	DFS	**0.6777**	**0.6765 ± 0.0006**	**0.6703 ± 0.0020**	500.0000 ± 0.0000
	BFT	0.6658	0.6590 ± 0.0035	0.6499 ± 0.0042	**500.0000 ± 0.0000**
Dna	RND	0.1966	0.1950 ± 0.0007	0.1942 ± 0.0007	482.3000 ± 57.9816
	FAS	0.1992	0.1982 ± 0.0004	0.1953 ± 0.0002	500.0000 ± 0.0000
	DFS	**0.1998**	**0.1991 ± 0.0003**	**0.1990 ± 0.0002**	**258.7000 ± 42.6580**
	BFT	0.1980	0.1965 ± 0.0006	0.1962 ± 0.0006	430.3300 ± 98.0880
Kosarek	RND	0.1772	0.1705 ± 0.0031	0.1667 ± 0.0031	500.0000 ± 0.0000
	FAS	0.1785	0.1762 ± 0.0008	0.1708 ± 0.0005	500.0000 ± 0.0000
	DFS	0.1777	0.1758 ± 0.0008	0.1666 ± 0.0005	500.0000 ± 0.0000
	BFT	**0.1822**	**0.1791 ± 0.0020**	**0.1773 ± 0.0024**	**495.5000 ± 31.6638**
Msweb	RND	0.1162	0.1123 ± 0.0020	0.1115 ± 0.0020	474.4000 ± 71.3265
	FAS	0.1198	**0.1185 ± 0.0006**	**0.1170 ± 0.0001**	**473.1800 ± 73.9972**
	DFS	0.1144	0.1125 ± 0.0009	0.1097 ± 0.0007	495.4900 ± 31.8759
	BFT	**0.1216**	0.1177 ± 0.0020	0.1165 ± 0.0018	479.6200 ± 63.0406
Book	RND	0.1212	0.1190 ± 0.0009	0.1189 ± 0.0009	261.9800 ± 19.4863
	FAS	**0.1265**	**0.1257 ± 0.0003**	**0.1257 ± 0.0002**	**215.7300 ± 16.7542**
	DFS	0.1194	0.1185 ± 0.0004	0.1184 ± 0.0004	225.8100 ± 15.5672
	BFT	0.1259	0.1250 ± 0.0005	0.1249 ± 0.0005	237.0800 ± 6.1294
Tmovie	RND	0.3155	0.3118 ± 0.0016	0.3113 ± 0.0016	125.2800 ± 4.1147
	FAS	0.3266	**0.3249 ± 0.0010**	**0.3247 ± 0.0010**	100.3100 ± 2.4192
	DFS	0.3103	0.3072 ± 0.0014	0.3065 ± 0.0015	192.0700 ± 11.5848
	BFT	**0.3267**	0.3234 ± 0.0012	0.3229 ± 0.0012	**94.6100 ± 1.8797**
Cwebkb	RND	0.1268	0.1240 ± 0.0012	0.1239 ± 0.0012	143.6600 ± 2.7014
	FAS	0.1337	**0.1332 ± 0.0002**	**0.1331 ± 0.0002**	138.1100 ± 1.6811
	DFS	0.1222	0.1202 ± 0.0009	0.1201 ± 0.0009	174.6600 ± 8.1380
	BFT	**0.1337**	0.1322 ± 0.0007	0.1321 ± 0.0007	**133.1900 ± 1.0982**
Cr52	RND	0.1706	0.1608 ± 0.0035	0.1606 ± 0.0034	113.2900 ± 3.4650
	FAS	**0.1792**	**0.1783 ± 0.0004**	**0.1781 ± 0.0004**	96.2800 ± 4.0802
	DFS	0.1500	0.1481 ± 0.0010	0.1475 ± 0.0010	144.6700 ± 6.6105
	BFT	0.1788	0.1749 ± 0.0022	0.1747 ± 0.0022	**91.9500 ± 3.6801**
C20ng	RND	0.0884	0.0860 ± 0.0011	0.0860 ± 0.0011	39.1500 ± 1.1044
	FAS	0.0942	**0.0938 ± 0.0002**	**0.0938 ± 0.0002**	37.4600 ± 2.5836
	DFS	0.0852	0.0842 ± 0.0004	0.0842 ± 0.0004	34.2500 ± 2.9418
	BFT	**0.0946**	0.0937 ± 0.0006	0.0937 ± 0.0006	**33.9600 ± 0.8399**
Bbc	RND	0.0751	0.0732 ± 0.0008	0.0731 ± 0.0008	191.5000 ± 2.8833
	FAS	**0.0802**	**0.0799 ± 0.0001**	**0.0798 ± 0.0001**	**172.2000 ± 16.5755**
	DFS	0.0744	0.0732 ± 0.0005	0.0732 ± 0.0005	206.0400 ± 12.6099
	BFT	0.0773	0.0760 ± 0.0005	0.0760 ± 0.0005	185.8700 ± 5.7746
Ad	RND	0.7106	0.7064 ± 0.0020	0.7063 ± 0.0020	**205.5200 ± 5.3494**
	FAS	0.7145	0.7109 ± 0.0018	0.7106 ± 0.0017	207.1200 ± 8.3428
	DFS	**0.7177**	**0.7140 ± 0.0017**	**0.7137 ± 0.0017**	208.5600 ± 6.6838
	BFT	0.7127	0.7093 ± 0.0019	0.7091 ± 0.0019	216.2600 ± 6.4159

Table C.2: *Best score obtained, Average best score obtained, Average initial score generated, Average number of iterations (Avg. It.) using greedy selection (best values in bold)*

Dataset	Approach	Best Score	Avg. Best Score	Avg. Initial Score	Avg. It.
Nltcs	RND	**0.3466**	0.3466 ± 0.0001	0.3432 ± 0.0009	**489.2400 ± 30.6090**
	FAS	0.3466	**0.3466 ± 0.0001**	0.3436 ± 0.0000	494.6700 ± 21.0368
	DFS	0.3466	0.3466 ± 0.0001	**0.3442 ± 0.0002**	492.2400 ± 23.5754
	BFT	0.3466	0.3466 ± 0.0001	0.3432 ± 0.0008	495.5800 ± 17.8655
Msnbc	RND	**0.0835**	**0.0819 ± 0.0011**	0.0750 ± 0.0035	475.5500 ± 52.4361
	FAS	0.0832	0.0808 ± 0.0007	**0.0771 ± 0.0000**	**424.6500 ± 88.6290**
	DFS	0.0835	0.0817 ± 0.0007	0.0764 ± 0.0013	464.1100 ± 67.7804
	BFT	0.0835	0.0817 ± 0.0011	0.0737 ± 0.0035	456.4600 ± 68.9987
Kdd	RND	0.1391	0.1353 ± 0.0020	0.1260 ± 0.0027	500.0000 ± 0.0000
	FAS	0.1398	**0.1372 ± 0.0008**	0.1230 ± 0.0005	500.0000 ± 0.0000
	DFS	0.1389	0.1368 ± 0.0011	0.1181 ± 0.0002	500.0000 ± 0.0000
	BFT	**0.1408**	0.1371 ± 0.0020	**0.1262 ± 0.0029**	**500.0000 ± 0.0000**
Plants	RND	0.5536	0.5460 ± 0.0037	0.5226 ± 0.0056	500.0000 ± 0.0000
	FAS	0.5467	0.5404 ± 0.0028	0.5209 ± 0.0017	500.0000 ± 0.0000
	DFS	**0.5572**	**0.5543 ± 0.0015**	**0.5352 ± 0.0018**	500.0000 ± 0.0000
	BFT	0.5521	0.5455 ± 0.0031	0.5255 ± 0.0044	**500.0000 ± 0.0000**
Baudio	RND	0.1447	0.1384 ± 0.0026	0.1287 ± 0.0030	500.0000 ± 0.0000
	FAS	0.1451	0.1428 ± 0.0009	0.1310 ± 0.0003	500.0000 ± 0.0000
	DFS	0.1477	**0.1447 ± 0.0014**	0.1221 ± 0.0011	500.0000 ± 0.0000
	BFT	**0.1478**	0.1434 ± 0.0022	**0.1376 ± 0.0025**	**500.0000 ± 0.0000**
Bnetflix	RND	0.0972	0.0937 ± 0.0018	0.0868 ± 0.0023	500.0000 ± 0.0000
	FAS	0.0995	0.0968 ± 0.0010	0.0872 ± 0.0002	500.0000 ± 0.0000
	DFS	**0.1002**	**0.0976 ± 0.0012**	0.0807 ± 0.0013	500.0000 ± 0.0000
	BFT	0.0981	0.0942 ± 0.0016	**0.0896 ± 0.0016**	**500.0000 ± 0.0000**
Jester	RND	0.1286	0.1238 ± 0.0026	0.1148 ± 0.0031	500.0000 ± 0.0000
	FAS	**0.1351**	**0.1313 ± 0.0013**	**0.1202 ± 0.0004**	500.0000 ± 0.0000
	DFS	0.1319	0.1288 ± 0.0017	0.1080 ± 0.0012	500.0000 ± 0.0000
	BFT	0.1317	0.1256 ± 0.0026	0.1189 ± 0.0024	**500.0000 ± 0.0000**
Accidents	RND	0.3290	0.3150 ± 0.0076	0.2990 ± 0.0083	500.0000 ± 0.0000
	FAS	0.3401	0.3366 ± 0.0018	**0.3205 ± 0.0007**	500.0000 ± 0.0000
	DFS	**0.3417**	**0.3381 ± 0.0027**	0.3045 ± 0.0012	500.0000 ± 0.0000
	BFT	0.3335	0.3153 ± 0.0059	0.3041 ± 0.0054	**500.0000 ± 0.0000**
Tretail	RND	0.0436	0.0415 ± 0.0019	0.0405 ± 0.0021	469.6200 ± 72.4220
	FAS	0.0441	**0.0437 ± 0.0001**	**0.0436 ± 0.0000**	**322.4500 ± 103.2411**
	DFS	0.0432	0.0426 ± 0.0004	0.0366 ± 0.0001	500.0000 ± 0.0000
	BFT	**0.0444**	0.0436 ± 0.0006	0.0433 ± 0.0006	435.7200 ± 92.3181
Pumsb_star	RND	0.6296	0.6161 ± 0.0063	0.6056 ± 0.0064	500.0000 ± 0.0000
	FAS	0.6443	0.6394 ± 0.0019	0.6302 ± 0.0013	500.0000 ± 0.0000
	DFS	**0.6475**	**0.6437 ± 0.0019**	**0.6331 ± 0.0023**	500.0000 ± 0.0000
	BFT	0.6239	0.6099 ± 0.0053	0.6008 ± 0.0051	**500.0000 ± 0.0000**
Dna	RND	0.1914	0.1894 ± 0.0008	0.1884 ± 0.0008	475.3300 ± 68.1595
	FAS	**0.1949**	**0.1946 ± 0.0001**	**0.1945 ± 0.0001**	**251.0700 ± 0.3258**
	DFS	0.1949	0.1938 ± 0.0006	0.1923 ± 0.0008	377.9200 ± 122.8961
	BFT	0.1921	0.1908 ± 0.0006	0.1904 ± 0.0006	449.1000 ± 88.2771
Kosarek	RND	0.1707	0.1623 ± 0.0045	0.1572 ± 0.0042	500.0000 ± 0.0000
	FAS	0.1771	**0.1751 ± 0.0009**	0.1698 ± 0.0008	500.0000 ± 0.0000
	DFS	0.1647	0.1620 ± 0.0012	0.1534 ± 0.0016	500.0000 ± 0.0000
	BFT	**0.1779**	0.1738 ± 0.0024	**0.1713 ± 0.0028**	**500.0000 ± 0.0000**
Msweb	RND	0.1171	0.1073 ± 0.0034	0.1059 ± 0.0033	**476.8100 ± 65.5329**
	FAS	0.1207	**0.1186 ± 0.0007**	**0.1166 ± 0.0000**	487.7900 ± 48.9247
	DFS	0.1102	0.1080 ± 0.0016	0.0943 ± 0.0026	500.0000 ± 0.0000
	BFT	**0.1209**	0.1154 ± 0.0027	0.1131 ± 0.0026	491.6800 ± 41.1648
Book	RND	0.1181	0.1151 ± 0.0012	0.1145 ± 0.0012	488.9500 ± 48.7480
	FAS	**0.1240**	**0.1229 ± 0.0004**	0.1211 ± 0.0003	495.6500 ± 30.6173
	DFS	0.1142	0.1129 ± 0.0007	0.1107 ± 0.0007	485.1800 ± 58.9566
	BFT	0.1237	0.1222 ± 0.0007	**0.1218 ± 0.0008**	**463.0800 ± 80.6353**
Tmovie	RND	0.2882	0.2789 ± 0.0038	0.2769 ± 0.0036	497.7400 ± 22.6000
	FAS	0.3013	0.2966 ± 0.0018	0.2876 ± 0.0006	500.0000 ± 0.0000
	DFS	0.2721	0.2655 ± 0.0034	0.2514 ± 0.0041	500.0000 ± 0.0000
	BFT	**0.3086**	**0.3000 ± 0.0036**	**0.2976 ± 0.0033**	**497.5100 ± 24.9000**
Cwebkb	RND	0.1211	0.1165 ± 0.0017	0.1162 ± 0.0017	407.4600 ± 56.5002
	FAS	0.1286	**0.1277 ± 0.0005**	0.1246 ± 0.0001	454.4100 ± 3.6903
	DFS	0.1145	0.1124 ± 0.0008	0.1100 ± 0.0007	443.6000 ± 15.7249
	BFT	**0.1287**	0.1268 ± 0.0011	**0.1265 ± 0.0011**	**379.3900 ± 45.1471**
Cr52	RND	0.1606	0.1447 ± 0.0055	0.1441 ± 0.0053	389.8400 ± 47.7448
	FAS	0.1667	**0.1656 ± 0.0006**	**0.1629 ± 0.0002**	390.8200 ± 20.8658
	DFS	0.1428	0.1346 ± 0.0028	0.1307 ± 0.0029	406.1400 ± 11.1591
	BFT	**0.1699**	0.1622 ± 0.0038	0.1616 ± 0.0037	**381.4200 ± 25.4845**
C20ng	RND	0.0760	0.0723 ± 0.0021	0.0720 ± 0.0021	263.9200 ± 10.5550
	FAS	0.0844	**0.0839 ± 0.0003**	0.0822 ± 0.0001	259.2600 ± 16.6161
	DFS	0.0635	0.0617 ± 0.0007	0.0595 ± 0.0007	271.0300 ± 3.5175
	BFT	**0.0874**	0.0838 ± 0.0018	**0.0836 ± 0.0017**	**248.4500 ± 5.6307**
Bbc	RND	0.0727	0.0709 ± 0.0009	0.0708 ± 0.0009	337.6100 ± 27.1162
	FAS	**0.0797**	**0.0791 ± 0.0002**	**0.0790 ± 0.0001**	289.5800 ± 45.5412
	DFS	0.0741	0.0715 ± 0.0011	0.0714 ± 0.0011	**256.6500 ± 18.4470**
	BFT	0.0759	0.0742 ± 0.0007	0.0741 ± 0.0007	335.4600 ± 37.8736
Ad	RND	0.6991	0.6935 ± 0.0025	0.6934 ± 0.0025	**213.4400 ± 8.6156**
	FAS	0.7051	**0.7017 ± 0.0015**	**0.7009 ± 0.0014**	215.1300 ± 6.7460
	DFS	**0.7065**	0.7009 ± 0.0020	0.7001 ± 0.0018	219.2700 ± 6.8133
	BFT	0.7030	0.6990 ± 0.0020	0.6988 ± 0.0020	241.9700 ± 26.0801

Table C.3: *Best score obtained, Average best score obtained, Average initial score generated, Average number of iterations (Avg. It.) using independence selection (best values in bold)*

Appendix D

Empirical Results: Tabu Search

Dataset	Approach	Best Score	Avg. Best Score	Avg. Initial Score	Avg. It.
Nltcs	RND	0.3456	0.3440 ± 0.0008	0.3431 ± 0.0010	**6.4400 ± 3.3794**
	FAS	0.3458	0.3454 ± 0.0002	0.3436 ± 0.0000	13.0200 ± 2.4204
	DFS	**0.3459**	**0.3459 ± 0.0000**	**0.3443 ± 0.0002**	11.3800 ± 1.1615
	BFT	0.3453	0.3441 ± 0.0007	0.3431 ± 0.0008	6.8500 ± 3.3676
Msnbc	RND	0.0827	0.0805 ± 0.0009	0.0791 ± 0.0011	11.9900 ± 5.0000
	FAS	0.0818	**0.0818 ± 0.0000**	**0.0817 ± 0.0000**	**4.0000 ± 0.0000**
	DFS	0.0802	0.0802 ± 0.0000	0.0788 ± 0.0003	22.9100 ± 0.7261
	BFT	**0.0827**	0.0813 ± 0.0009	0.0802 ± 0.0011	8.9200 ± 3.7784
Kdd	RND	0.1423	0.1406 ± 0.0010	0.1402 ± 0.0009	**7.3000 ± 3.0368**
	FAS	**0.1452**	**0.1449 ± 0.0001**	**0.1440 ± 0.0001**	16.1600 ± 2.6656
	DFS	0.1441	0.1437 ± 0.0003	0.1427 ± 0.0005	23.3600 ± 2.6112
	BFT	0.1446	0.1427 ± 0.0009	0.1423 ± 0.0009	8.3700 ± 3.6698
Plants	RND	0.5737	0.5717 ± 0.0009	0.5711 ± 0.0010	8.6600 ± 3.3461
	FAS	0.5749	0.5742 ± 0.0004	0.5725 ± 0.0004	21.6800 ± 3.6206
	DFS	0.5761	**0.5747 ± 0.0008**	0.5710 ± 0.0006	36.3200 ± 4.0123
	BFT	**0.5768**	0.5743 ± 0.0009	**0.5738 ± 0.0009**	**7.5600 ± 3.1053**
Baudio	RND	0.1562	0.1539 ± 0.0010	0.1534 ± 0.0010	11.5200 ± 6.3029
	FAS	**0.1604**	**0.1599 ± 0.0002**	**0.1588 ± 0.0002**	25.4200 ± 3.7259
	DFS	0.1587	0.1580 ± 0.0003	0.1564 ± 0.0003	35.9600 ± 4.7904
	BFT	0.1593	0.1573 ± 0.0008	0.1569 ± 0.0008	**9.8300 ± 4.0228**
Bnetflix	RND	0.1038	0.1014 ± 0.0009	0.1011 ± 0.0008	9.9300 ± 3.7933
	FAS	**0.1078**	**0.1073 ± 0.0002**	**0.1056 ± 0.0002**	33.2100 ± 5.0338
	DFS	0.1065	0.1055 ± 0.0005	0.1042 ± 0.0006	28.8000 ± 4.6537
	BFT	0.1025	0.1009 ± 0.0007	0.1007 ± 0.0007	**9.3600 ± 3.3621**
Jester	RND	0.1423	0.1405 ± 0.0010	0.1401 ± 0.0010	12.6200 ± 5.6974
	FAS	**0.1467**	**0.1457 ± 0.0005**	**0.1445 ± 0.0004**	28.9400 ± 3.7921
	DFS	0.1457	0.1448 ± 0.0005	0.1438 ± 0.0006	30.7100 ± 4.1955
	BFT	0.1421	0.1404 ± 0.0008	0.1401 ± 0.0008	**12.3400 ± 3.9215**
Accidents	RND	0.3623	0.3535 ± 0.0043	0.3527 ± 0.0042	**8.7300 ± 3.6704**
	FAS	0.3721	0.3699 ± 0.0009	**0.3671 ± 0.0007**	28.1200 ± 4.7296
	DFS	**0.3740**	**0.3734 ± 0.0003**	0.3665 ± 0.0011	35.7800 ± 3.2085
	BFT	0.3651	0.3569 ± 0.0037	0.3562 ± 0.0037	10.3900 ± 4.1655
Tretail	RND	0.0433	0.0407 ± 0.0020	0.0406 ± 0.0020	**2.5500 ± 1.6476**
	FAS	0.0443	**0.0441 ± 0.0001**	**0.0438 ± 0.0000**	15.8800 ± 3.1791
	DFS	0.0430	0.0425 ± 0.0001	0.0406 ± 0.0001	31.5600 ± 3.4970
	BFT	**0.0446**	0.0437 ± 0.0006	0.0437 ± 0.0006	3.2000 ± 1.8910
Pumsb_star	RND	0.6875	0.6816 ± 0.0024	0.6809 ± 0.0024	**9.1400 ± 4.2806**
	FAS	0.6914	0.6892 ± 0.0009	0.6864 ± 0.0010	16.9700 ± 2.7431
	DFS	**0.6931**	**0.6924 ± 0.0003**	**0.6903 ± 0.0005**	13.7300 ± 1.4692
	BFT	0.6802	0.6738 ± 0.0027	0.6731 ± 0.0026	9.7900 ± 3.7774
Dna	RND	0.1969	0.1958 ± 0.0005	0.1958 ± 0.0005	6.6300 ± 6.3543
	FAS	0.1989	0.1984 ± 0.0003	0.1960 ± 0.0002	59.0500 ± 9.4499
	DFS	**0.2005**	**0.2002 ± 0.0001**	**0.1990 ± 0.0003**	48.7200 ± 2.9918
	BFT	0.1979	0.1971 ± 0.0004	0.1970 ± 0.0004	**5.4500 ± 6.7992**
Kosarek	RND	0.1752	0.1665 ± 0.0039	0.1658 ± 0.0037	16.1200 ± 7.9598
	FAS	**0.1822**	**0.1811 ± 0.0006**	**0.1791 ± 0.0006**	30.3900 ± 6.0668
	DFS	0.1808	0.1786 ± 0.0012	0.1766 ± 0.0015	43.1800 ± 4.6740
	BFT	0.1821	0.1790 ± 0.0022	0.1786 ± 0.0024	**10.7600 ± 7.6820**
Msweb	RND	0.1164	0.1113 ± 0.0023	0.1111 ± 0.0022	**2.6800 ± 1.7803**
	FAS	**0.1227**	**0.1218 ± 0.0004**	**0.1203 ± 0.0001**	23.1000 ± 2.7979
	DFS	0.1149	0.1128 ± 0.0010	0.1110 ± 0.0009	28.0500 ± 4.2029
	BFT	0.1222	0.1175 ± 0.0025	0.1172 ± 0.0025	4.3600 ± 2.1345
Book	RND	0.1161	0.1144 ± 0.0008	0.1143 ± 0.0008	4.3500 ± 0.5000
	FAS	**0.1222**	**0.1215 ± 0.0003**	**0.1213 ± 0.0003**	3.9900 ± 0.1738
	DFS	0.1199	0.1187 ± 0.0006	0.1185 ± 0.0006	4.0700 ± 0.8675
	BFT	0.1218	0.1203 ± 0.0005	0.1202 ± 0.0005	**3.6900 ± 0.8726**
Tmovie	RND	0.2886	0.2853 ± 0.0014	0.2851 ± 0.0014	3.9000 ± 0.3333
	FAS	**0.3038**	**0.3022 ± 0.0007**	**0.3018 ± 0.0007**	3.0300 ± 0.1714
	DFS	0.2967	0.2904 ± 0.0025	0.2898 ± 0.0024	4.0400 ± 0.1969
	BFT	0.2984	0.2953 ± 0.0013	0.2950 ± 0.0013	**2.8900 ± 0.3145**
Cwebkb	RND	0.1123	0.1091 ± 0.0014	0.1090 ± 0.0013	2.8300 ± 0.3775
	FAS	**0.1233**	**0.1222 ± 0.0005**	**0.1220 ± 0.0005**	3.0000 ± 0.0000
	DFS	0.1212	0.1196 ± 0.0008	0.1194 ± 0.0008	2.7600 ± 0.4292
	BFT	0.1196	0.1176 ± 0.0009	0.1175 ± 0.0009	**2.1000 ± 0.3624**
Cr52	RND	0.1431	0.1341 ± 0.0040	0.1340 ± 0.0039	2.8300 ± 0.3775
	FAS	**0.1593**	**0.1573 ± 0.0010**	**0.1570 ± 0.0010**	2.0900 ± 0.2876
	DFS	0.1577	0.1560 ± 0.0007	0.1557 ± 0.0007	2.0000 ± 0.0000
	BFT	0.1548	0.1488 ± 0.0031	0.1486 ± 0.0031	**1.9500 ± 0.2973**
C20ng	RND	0.0701	0.0680 ± 0.0010	0.0680 ± 0.0009	1.0000 ± 0.0000
	FAS	**0.0770**	**0.0764 ± 0.0003**	**0.0763 ± 0.0003**	1.0000 ± 0.0000
	DFS	0.0753	0.0742 ± 0.0005	0.0741 ± 0.0005	1.0000 ± 0.0000
	BFT	0.0752	0.0734 ± 0.0006	0.0734 ± 0.0006	**1.0000 ± 0.0000**
Bbc	RND	0.0671	0.0649 ± 0.0009	0.0649 ± 0.0009	7.1600 ± 1.4405
	FAS	**0.0770**	**0.0765 ± 0.0003**	**0.0762 ± 0.0003**	**6.9400 ± 0.2778**
	DFS	0.0735	0.0718 ± 0.0008	0.0715 ± 0.0008	7.1700 ± 1.5574
	BFT	0.0729	0.0718 ± 0.0005	0.0718 ± 0.0005	7.2700 ± 1.3170
Ad	RND	0.6891	0.6853 ± 0.0021	0.6852 ± 0.0021	**1.3600 ± 1.2433**
	FAS	0.7002	0.6960 ± 0.0019	0.6950 ± 0.0019	5.9700 ± 0.4370
	DFS	**0.7006**	**0.6967 ± 0.0019**	**0.6956 ± 0.0018**	5.2200 ± 0.5041
	BFT	0.6967	0.6919 ± 0.0026	0.6918 ± 0.0025	2.5200 ± 1.7025

Table D.1: *Best score obtained, Average best score obtained, Average initial score generated, Average number of iterations (Avg. It.) using sequential selection (best values in bold)*

Dataset	Approach	Best Score	Avg. Best Score	Avg. Initial Score	Avg. It.
Nltcs	RND	0.3416	0.3356 ± 0.0029	0.3326 ± 0.0031	5.2700 ± 2.5657
	FAS	**0.3416**	**0.3411 ± 0.0005**	**0.3377 ± 0.0010**	**4.9000 ± 1.4599**
	DFS	0.3410	0.3408 ± 0.0001	0.3373 ± 0.0007	5.4400 ± 0.5915
	BFT	0.3403	0.3354 ± 0.0024	0.3319 ± 0.0025	6.3600 ± 2.9798
Msnbc	RND	**0.0585**	0.0523 ± 0.0029	0.0492 ± 0.0032	8.2400 ± 3.6073
	FAS	0.0567	**0.0565 ± 0.0002**	0.0540 ± 0.0000	4.8000 ± 0.5860
	DFS	0.0553	0.0549 ± 0.0004	**0.0543 ± 0.0005**	**2.2800 ± 0.4513**
	BFT	0.0582	0.0542 ± 0.0025	0.0520 ± 0.0031	6.1900 ± 3.0672
Kdd	RND	0.1382	0.1361 ± 0.0011	0.1355 ± 0.0012	7.5000 ± 3.5971
	FAS	**0.1405**	**0.1396 ± 0.0005**	**0.1388 ± 0.0002**	11.6500 ± 3.9348
	DFS	0.1388	0.1380 ± 0.0003	0.1371 ± 0.0002	26.1100 ± 2.7076
	BFT	0.1404	0.1378 ± 0.0011	0.1371 ± 0.0011	**7.2900 ± 2.6716**
Plants	RND	0.5627	0.5585 ± 0.0020	0.5571 ± 0.0020	8.8800 ± 2.8790
	FAS	0.5653	**0.5640 ± 0.0006**	**0.5616 ± 0.0006**	11.5500 ± 2.6029
	DFS	**0.5654**	0.5634 ± 0.0007	0.5574 ± 0.0009	30.9400 ± 3.3419
	BFT	0.5649	0.5603 ± 0.0020	0.5586 ± 0.0020	**8.7200 ± 3.2289**
Baudio	RND	0.1570	0.1543 ± 0.0010	0.1537 ± 0.0010	10.4200 ± 3.1883
	FAS	0.1599	0.1585 ± 0.0005	0.1565 ± 0.0005	25.1000 ± 4.0614
	DFS	**0.1609**	**0.1600 ± 0.0004**	**0.1575 ± 0.0003**	34.6600 ± 4.1760
	BFT	0.1589	0.1573 ± 0.0007	0.1568 ± 0.0007	**8.9200 ± 3.6367**
Bnetflix	RND	0.1033	0.1011 ± 0.0008	0.1007 ± 0.0008	10.5700 ± 3.5341
	FAS	0.1046	0.1037 ± 0.0004	0.1021 ± 0.0004	24.8500 ± 3.8727
	DFS	**0.1062**	**0.1054 ± 0.0004**	**0.1038 ± 0.0005**	31.1000 ± 4.3193
	BFT	0.1023	0.1006 ± 0.0007	0.1003 ± 0.0006	**9.2100 ± 3.7126**
Jester	RND	0.1424	0.1400 ± 0.0010	0.1394 ± 0.0010	12.4000 ± 5.1109
	FAS	**0.1465**	**0.1458 ± 0.0003**	**0.1446 ± 0.0003**	20.9200 ± 3.4922
	DFS	0.1464	0.1454 ± 0.0004	0.1435 ± 0.0005	43.8800 ± 4.8706
	BFT	0.1421	0.1404 ± 0.0008	0.1399 ± 0.0008	**12.2500 ± 6.3060**
Accidents	RND	0.3747	0.3652 ± 0.0043	0.3640 ± 0.0044	**8.2900 ± 3.7289**
	FAS	0.3805	0.3752 ± 0.0018	0.3709 ± 0.0014	25.7900 ± 3.6936
	DFS	**0.3835**	**0.3817 ± 0.0008**	**0.3734 ± 0.0005**	42.6900 ± 3.2526
	BFT	0.3735	0.3654 ± 0.0038	0.3645 ± 0.0038	9.0700 ± 3.5683
Tretail	RND	0.0434	0.0420 ± 0.0008	0.0420 ± 0.0008	**2.8200 ± 1.6780**
	FAS	0.0443	**0.0440 ± 0.0001**	0.0430 ± 0.0001	16.2800 ± 3.0354
	DFS	0.0437	0.0435 ± 0.0002	0.0424 ± 0.0002	29.5700 ± 3.3067
	BFT	**0.0444**	0.0435 ± 0.0006	**0.0435 ± 0.0006**	3.3800 ± 2.0830
Pumsb_star	RND	0.6638	0.6527 ± 0.0052	0.6514 ± 0.0054	7.6300 ± 3.6004
	FAS	0.6745	0.6684 ± 0.0020	0.6663 ± 0.0017	21.9700 ± 4.0488
	DFS	**0.6763**	**0.6734 ± 0.0019**	**0.6703 ± 0.0019**	25.3100 ± 3.8735
	BFT	0.6610	0.6506 ± 0.0041	0.6496 ± 0.0040	**7.4400 ± 3.6134**
Dna	RND	0.1958	0.1942 ± 0.0007	0.1940 ± 0.0007	**5.5600 ± 3.8725**
	FAS	0.1986	0.1978 ± 0.0003	0.1953 ± 0.0002	70.5700 ± 23.0517
	DFS	**0.2009**	**0.2007 ± 0.0001**	**0.1990 ± 0.0002**	56.0400 ± 5.4826
	BFT	0.1975	0.1963 ± 0.0006	0.1962 ± 0.0006	6.1000 ± 5.7884
Kosarek	RND	0.1737	0.1671 ± 0.0032	0.1665 ± 0.0031	17.4000 ± 14.1450
	FAS	0.1760	0.1742 ± 0.0008	0.1707 ± 0.0005	89.5200 ± 54.8593
	DFS	0.1718	0.1705 ± 0.0006	0.1667 ± 0.0006	73.3000 ± 20.3854
	BFT	**0.1816**	**0.1780 ± 0.0023**	**0.1776 ± 0.0023**	**11.3500 ± 7.0114**
Msweb	RND	0.1159	0.1113 ± 0.0022	0.1111 ± 0.0022	**2.3300 ± 1.5895**
	FAS	0.1194	**0.1186 ± 0.0003**	**0.1170 ± 0.0001**	27.8400 ± 2.9771
	DFS	0.1132	0.1117 ± 0.0006	0.1096 ± 0.0008	36.6600 ± 4.1175
	BFT	**0.1218**	0.1171 ± 0.0021	0.1168 ± 0.0020	4.2300 ± 2.2422
Book	RND	0.1206	0.1188 ± 0.0008	0.1187 ± 0.0008	9.9800 ± 3.8032
	FAS	**0.1270**	**0.1264 ± 0.0002**	**0.1257 ± 0.0002**	24.8100 ± 0.8372
	DFS	0.1198	0.1190 ± 0.0004	0.1184 ± 0.0004	17.9200 ± 2.7254
	BFT	0.1262	0.1251 ± 0.0005	0.1250 ± 0.0005	**7.9700 ± 2.6532**
Tmovie	RND	0.3159	0.3116 ± 0.0018	0.3112 ± 0.0018	8.4100 ± 1.4361
	FAS	**0.3271**	**0.3254 ± 0.0009**	**0.3245 ± 0.0009**	7.7400 ± 0.9705
	DFS	0.3106	0.3077 ± 0.0014	0.3065 ± 0.0015	16.4600 ± 1.1927
	BFT	0.3259	0.3238 ± 0.0010	0.3233 ± 0.0010	**7.1400 ± 1.4976**
Cwebkb	RND	0.1269	0.1242 ± 0.0011	0.1241 ± 0.0010	6.9500 ± 1.0384
	FAS	**0.1340**	**0.1334 ± 0.0003**	**0.1332 ± 0.0002**	6.5800 ± 1.5517
	DFS	0.1226	0.1204 ± 0.0008	0.1201 ± 0.0008	6.8200 ± 1.8278
	BFT	0.1339	0.1324 ± 0.0007	0.1323 ± 0.0007	**6.2700 ± 1.2781**
Cr52	RND	0.1672	0.1607 ± 0.0031	0.1605 ± 0.0031	5.5900 ± 0.6528
	FAS	**0.1798**	**0.1787 ± 0.0005**	**0.1781 ± 0.0004**	4.9500 ± 0.2611
	DFS	0.1502	0.1485 ± 0.0009	0.1476 ± 0.0009	6.4200 ± 0.6694
	BFT	0.1787	0.1749 ± 0.0022	0.1747 ± 0.0022	**3.8000 ± 0.7785**
C20ng	RND	0.0883	0.0860 ± 0.0011	0.0859 ± 0.0011	1.7900 ± 0.4094
	FAS	0.0943	**0.0939 ± 0.0002**	**0.0938 ± 0.0002**	**1.7800 ± 0.4163**
	DFS	0.0855	0.0844 ± 0.0005	0.0843 ± 0.0005	2.1600 ± 0.3685
	BFT	**0.0951**	0.0938 ± 0.0007	0.0937 ± 0.0006	2.0000 ± 0.0000
Bbc	RND	0.0751	0.0731 ± 0.0008	0.0730 ± 0.0008	**6.0600 ± 1.5296**
	FAS	**0.0806**	**0.0802 ± 0.0002**	**0.0799 ± 0.0001**	7.1500 ± 1.4097
	DFS	0.0752	0.0736 ± 0.0006	0.0733 ± 0.0006	8.7700 ± 0.6333
	BFT	0.0770	0.0760 ± 0.0005	0.0760 ± 0.0005	7.5200 ± 1.0394
Ad	RND	0.7115	0.7065 ± 0.0019	0.7064 ± 0.0018	**1.5100 ± 1.2431**
	FAS	0.7149	0.7115 ± 0.0017	0.7106 ± 0.0017	5.6300 ± 0.4852
	DFS	**0.7189**	**0.7146 ± 0.0015**	**0.7138 ± 0.0015**	5.7900 ± 0.6860
	BFT	0.7130	0.7096 ± 0.0018	0.7095 ± 0.0017	2.9600 ± 1.6873

Table D.2: *Best score obtained, Average best score obtained, Average initial score generated, Average number of iterations (Avg. It.) using greedy selection (best values in bold)*

Dataset	Approach	Best Score	Avg. Best Score	Avg. Initial Score	Avg. It.
Nltcs	RND	0.3453	0.3439 ± 0.0007	0.3431 ± 0.0008	**6.0300 ± 3.2736**
	FAS	0.3458	0.3453 ± 0.0002	0.3436 ± 0.0000	11.9400 ± 2.1217
	DFS	**0.3459**	**0.3459 ± 0.0000**	**0.3443 ± 0.0002**	11.3800 ± 1.1615
	BFT	0.3453	0.3441 ± 0.0007	0.3431 ± 0.0008	6.4700 ± 3.4479
Msnbc	RND	**0.0815**	0.0790 ± 0.0019	0.0756 ± 0.0036	13.3300 ± 4.7163
	FAS	0.0798	0.0791 ± 0.0009	**0.0771 ± 0.0000**	**8.4900 ± 1.5986**
	DFS	0.0803	**0.0794 ± 0.0004**	0.0762 ± 0.0013	18.0700 ± 2.9653
	BFT	0.0813	0.0781 ± 0.0022	0.0732 ± 0.0036	12.1200 ± 5.0518
Kdd	RND	**0.1328**	0.1267 ± 0.0027	0.1256 ± 0.0029	**5.5000 ± 2.5126**
	FAS	0.1318	**0.1286 ± 0.0016**	0.1231 ± 0.0005	15.1900 ± 2.9049
	DFS	0.1272	0.1257 ± 0.0008	0.1181 ± 0.0002	29.0700 ± 4.0408
	BFT	0.1325	0.1275 ± 0.0027	**0.1262 ± 0.0027**	5.9100 ± 2.2967
Plants	RND	0.5356	0.5265 ± 0.0047	0.5235 ± 0.0050	7.1100 ± 3.3360
	FAS	0.5314	0.5280 ± 0.0019	0.5212 ± 0.0019	18.8800 ± 3.6798
	DFS	**0.5478**	**0.5467 ± 0.0007**	**0.5354 ± 0.0017**	37.5500 ± 2.3328
	BFT	0.5377	0.5279 ± 0.0042	0.5258 ± 0.0043	**6.9700 ± 3.3256**
Baudio	RND	0.1360	0.1304 ± 0.0028	0.1295 ± 0.0028	6.8900 ± 5.6406
	FAS	0.1408	**0.1387 ± 0.0010**	0.1310 ± 0.0003	27.0100 ± 2.8972
	DFS	0.1340	0.1316 ± 0.0010	0.1220 ± 0.0010	32.9700 ± 2.7651
	BFT	**0.1438**	0.1381 ± 0.0026	**0.1375 ± 0.0026**	**5.4900 ± 2.6874**
Bnetflix	RND	0.0919	0.0873 ± 0.0025	0.0867 ± 0.0024	6.0500 ± 2.7206
	FAS	**0.0930**	**0.0916 ± 0.0005**	0.0872 ± 0.0002	26.6900 ± 3.3806
	DFS	0.0910	0.0883 ± 0.0011	0.0807 ± 0.0012	43.1100 ± 6.1969
	BFT	0.0928	0.0899 ± 0.0016	**0.0896 ± 0.0016**	**5.2600 ± 3.4071**
Jester	RND	0.1234	0.1153 ± 0.0033	0.1145 ± 0.0032	**9.3200 ± 5.4493**
	FAS	**0.1282**	**0.1256 ± 0.0011**	**0.1201 ± 0.0004**	23.1700 ± 4.3787
	DFS	0.1208	0.1172 ± 0.0018	0.1079 ± 0.0013	43.0800 ± 4.8214
	BFT	0.1252	0.1199 ± 0.0025	0.1193 ± 0.0025	11.4000 ± 7.9760
Accidents	RND	0.3199	0.2995 ± 0.0077	0.2979 ± 0.0074	6.9800 ± 6.4244
	FAS	0.3321	**0.3258 ± 0.0025**	**0.3205 ± 0.0007**	16.8300 ± 3.2847
	DFS	**0.3326**	0.3227 ± 0.0030	0.3045 ± 0.0011	35.5500 ± 4.7213
	BFT	0.3207	0.3058 ± 0.0066	0.3049 ± 0.0064	**6.1900 ± 5.6563**
Tretail	RND	0.0436	0.0406 ± 0.0019	0.0405 ± 0.0019	**3.1300 ± 2.6729**
	FAS	0.0441	**0.0439 ± 0.0001**	**0.0436 ± 0.0000**	15.8500 ± 2.4137
	DFS	0.0427	0.0424 ± 0.0001	0.0366 ± 0.0001	33.2800 ± 3.5393
	BFT	**0.0443**	0.0434 ± 0.0010	0.0433 ± 0.0010	3.4100 ± 2.0846
Pumsb_star	RND	0.6245	0.6070 ± 0.0068	0.6053 ± 0.0066	5.7800 ± 2.9834
	FAS	0.6395	0.6354 ± 0.0021	0.6300 ± 0.0016	11.5600 ± 2.1664
	DFS	**0.6431**	**0.6382 ± 0.0022**	**0.6334 ± 0.0023**	17.8800 ± 2.8472
	BFT	0.6138	0.6010 ± 0.0053	0.6000 ± 0.0051	**5.7300 ± 3.0907**
Dna	RND	0.1905	0.1884 ± 0.0008	0.1883 ± 0.0008	3.9700 ± 4.1009
	FAS	0.1963	**0.1960 ± 0.0001**	**0.1945 ± 0.0001**	54.8900 ± 15.3412
	DFS	**0.1963**	0.1957 ± 0.0003	0.1924 ± 0.0008	32.6800 ± 6.1297
	BFT	0.1918	0.1906 ± 0.0005	0.1906 ± 0.0005	**3.5200 ± 4.9410**
Kosarek	RND	0.1702	0.1585 ± 0.0048	0.1576 ± 0.0046	22.2200 ± 28.1043
	FAS	0.1744	**0.1728 ± 0.0008**	0.1698 ± 0.0008	53.1500 ± 30.8464
	DFS	0.1600	0.1575 ± 0.0012	0.1533 ± 0.0014	81.0700 ± 22.3816
	BFT	**0.1767**	0.1719 ± 0.0031	**0.1714 ± 0.0031**	**10.8400 ± 12.3196**
Msweb	RND	0.1134	0.1060 ± 0.0031	0.1057 ± 0.0030	**2.9300 ± 1.8491**
	FAS	0.1198	**0.1184 ± 0.0005**	**0.1166 ± 0.0000**	21.6100 ± 3.3481
	DFS	0.1068	0.1016 ± 0.0019	0.0942 ± 0.0026	38.5500 ± 5.8713
	BFT	**0.1198**	0.1138 ± 0.0029	0.1132 ± 0.0029	3.8600 ± 2.1743
Book	RND	0.1173	0.1145 ± 0.0011	0.1144 ± 0.0011	8.1400 ± 3.4024
	FAS	**0.1239**	**0.1232 ± 0.0003**	0.1211 ± 0.0003	77.6700 ± 8.3908
	DFS	0.1145	0.1131 ± 0.0006	0.1107 ± 0.0007	85.5900 ± 7.9875
	BFT	0.1235	0.1219 ± 0.0007	**0.1217 ± 0.0007**	**5.9400 ± 2.7955**
Tmovie	RND	0.2853	0.2770 ± 0.0036	0.2764 ± 0.0035	22.8400 ± 23.6454
	FAS	0.2987	0.2955 ± 0.0011	0.2877 ± 0.0005	55.5100 ± 7.3478
	DFS	0.2665	0.2583 ± 0.0034	0.2509 ± 0.0037	78.6300 ± 7.9667
	BFT	**0.3068**	**0.2980 ± 0.0040**	**0.2973 ± 0.0039**	**15.9800 ± 14.9639**
Cwebkb	RND	0.1205	0.1165 ± 0.0018	0.1163 ± 0.0017	9.1000 ± 4.4845
	FAS	0.1273	**0.1268 ± 0.0002**	0.1247 ± 0.0001	20.7800 ± 1.5081
	DFS	0.1135	0.1117 ± 0.0008	0.1102 ± 0.0008	17.8100 ± 2.7696
	BFT	**0.1298**	0.1267 ± 0.0014	**0.1265 ± 0.0013**	**5.7900 ± 2.3838**
Cr52	RND	0.1544	0.1443 ± 0.0051	0.1441 ± 0.0051	8.9700 ± 3.6500
	FAS	0.1659	**0.1649 ± 0.0004**	**0.1629 ± 0.0003**	16.8200 ± 2.1291
	DFS	0.1442	0.1340 ± 0.0031	0.1312 ± 0.0031	20.0600 ± 0.9301
	BFT	**0.1690**	0.1629 ± 0.0030	0.1626 ± 0.0030	**5.1100 ± 2.3566**
C20ng	RND	0.0765	0.0719 ± 0.0021	0.0717 ± 0.0021	8.2300 ± 2.5260
	FAS	0.0836	0.0831 ± 0.0002	0.0823 ± 0.0001	9.6700 ± 2.8358
	DFS	0.0619	0.0604 ± 0.0008	0.0594 ± 0.0008	12.9900 ± 0.1000
	BFT	**0.0877**	**0.0834 ± 0.0019**	**0.0833 ± 0.0018**	**5.8100 ± 2.8664**
Bbc	RND	0.0727	0.0709 ± 0.0008	0.0708 ± 0.0008	**10.6300 ± 3.5295**
	FAS	**0.0801**	**0.0796 ± 0.0002**	**0.0789 ± 0.0002**	11.7200 ± 0.9648
	DFS	0.0740	0.0716 ± 0.0011	0.0711 ± 0.0011	12.1300 ± 1.9471
	BFT	0.0757	0.0741 ± 0.0007	0.0741 ± 0.0006	10.8900 ± 3.6678
Ad	RND	0.6987	0.6935 ± 0.0025	0.6933 ± 0.0024	**1.3600 ± 1.1148**
	FAS	0.7064	**0.7035 ± 0.0015**	**0.7014 ± 0.0014**	6.3100 ± 0.8610
	DFS	**0.7067**	0.7022 ± 0.0017	0.7001 ± 0.0018	6.2400 ± 0.9653
	BFT	0.7039	0.6986 ± 0.0022	0.6984 ± 0.0022	2.7200 ± 1.8591

Table D.3: *Best score obtained, Average best score obtained, Average initial score generated, Average number of iterations (Avg. It.) using independence selection (best values in bold)*

Appendix E

Empirical Results: Beam Search

Dataset	Approach	Best Score	Avg. Best Score	Avg. Initial Score	Avg. It.
Nltcs	RND	0.3455	0.3440 ± 0.0008	0.3431 ± 0.0009	**8.4700 ± 4.0488**
	FAS	0.3455	0.3452 ± 0.0003	0.3436 ± 0.0000	18.3100 ± 48.7087
	DFS	**0.3459**	**0.3459 ± 0.0000**	**0.3443 ± 0.0002**	19.0000 ± 1.5505
	BFT	0.3456	0.3441 ± 0.0008	0.3431 ± 0.0010	43.8000 ± 125.8486
Msnbc	RND	0.0823	0.0797 ± 0.0012	0.0789 ± 0.0009	218.4100 ± 240.8731
	FAS	0.0818	**0.0818 ± 0.0000**	**0.0817 ± 0.0000**	**5.0000 ± 0.0000**
	DFS	0.0801	0.0801 ± 0.0001	0.0788 ± 0.0003	29.9800 ± 0.7245
	BFT	**0.0825**	0.0808 ± 0.0012	0.0804 ± 0.0012	313.8600 ± 238.9730
Kdd	RND	0.1421	0.1407 ± 0.0008	0.1405 ± 0.0008	92.0700 ± 185.5769
	FAS	**0.1452**	**0.1445 ± 0.0004**	**0.1440 ± 0.0001**	272.8400 ± 241.5060
	DFS	0.1442	0.1436 ± 0.0005	0.1426 ± 0.0004	**89.0300 ± 152.5719**
	BFT	0.1441	0.1424 ± 0.0009	0.1422 ± 0.0009	239.3500 ± 246.6985
Plants	RND	0.5736	0.5716 ± 0.0010	0.5710 ± 0.0009	11.3600 ± 4.3147
	FAS	0.5746	0.5731 ± 0.0005	0.5726 ± 0.0003	15.0100 ± 1.8775
	DFS	0.5725	0.5711 ± 0.0007	0.5710 ± 0.0006	9.8600 ± 1.3259
	BFT	**0.5764**	**0.5743 ± 0.0008**	**0.5737 ± 0.0008**	**9.4400 ± 3.5854**
Baudio	RND	0.1567	0.1538 ± 0.0012	0.1535 ± 0.0011	12.1600 ± 3.1932
	FAS	**0.1594**	**0.1590 ± 0.0002**	**0.1588 ± 0.0002**	11.0100 ± 1.6545
	DFS	0.1574	0.1566 ± 0.0004	0.1565 ± 0.0004	**9.3700 ± 1.2606**
	BFT	0.1588	0.1571 ± 0.0008	0.1568 ± 0.0008	10.7900 ± 3.7478
Bnetflix	RND	0.1033	0.1015 ± 0.0009	0.1012 ± 0.0009	13.7500 ± 7.8758
	FAS	**0.1072**	**0.1063 ± 0.0004**	**0.1057 ± 0.0002**	22.0300 ± 2.6341
	DFS	0.1062	0.1049 ± 0.0006	0.1042 ± 0.0006	23.0200 ± 3.0384
	BFT	0.1026	0.1009 ± 0.0008	0.1006 ± 0.0007	**12.1700 ± 7.6357**
Jester	RND	0.1426	0.1403 ± 0.0009	0.1402 ± 0.0009	7.8400 ± 1.0222
	FAS	**0.1453**	**0.1446 ± 0.0004**	**0.1445 ± 0.0004**	**5.3300 ± 0.5136**
	DFS	0.1450	0.1439 ± 0.0006	0.1438 ± 0.0006	5.7200 ± 0.7924
	BFT	0.1416	0.1402 ± 0.0007	0.1400 ± 0.0007	7.6700 ± 0.9955
Accidents	RND	0.3624	0.3533 ± 0.0044	0.3525 ± 0.0044	**11.3100 ± 8.2679**
	FAS	0.3724	0.3696 ± 0.0012	**0.3670 ± 0.0007**	39.4800 ± 27.0213
	DFS	**0.3739**	**0.3732 ± 0.0003**	0.3664 ± 0.0009	43.7600 ± 3.8642
	BFT	0.3648	0.3575 ± 0.0035	0.3568 ± 0.0034	25.1400 ± 30.9317
Tretail	RND	0.0440	0.0407 ± 0.0023	0.0406 ± 0.0023	**3.5600 ± 1.4656**
	FAS	0.0443	**0.0441 ± 0.0001**	**0.0438 ± 0.0000**	17.5100 ± 3.0765
	DFS	0.0429	0.0426 ± 0.0001	0.0406 ± 0.0001	33.4600 ± 3.8858
	BFT	**0.0445**	0.0436 ± 0.0007	0.0435 ± 0.0007	3.9000 ± 1.9149
Pumsb_star	RND	0.6866	0.6812 ± 0.0022	0.6810 ± 0.0022	4.2100 ± 0.5911
	FAS	0.6893	0.6864 ± 0.0010	0.6864 ± 0.0010	2.9400 ± 0.2387
	DFS	**0.6915**	**0.6903 ± 0.0005**	**0.6902 ± 0.0005**	**2.8800 ± 0.3266**
	BFT	0.6787	0.6725 ± 0.0027	0.6723 ± 0.0026	4.6300 ± 0.5801
Dna	RND	0.1971	0.1958 ± 0.0006	0.1958 ± 0.0005	307.4500 ± 242.0314
	FAS	0.1985	0.1961 ± 0.0005	0.1960 ± 0.0002	486.9100 ± 74.8085
	DFS	**0.2004**	**0.1997 ± 0.0004**	**0.1990 ± 0.0003**	**297.2400 ± 225.3047**
	BFT	0.1977	0.1971 ± 0.0004	0.1970 ± 0.0004	331.6200 ± 235.7827
Kosarek	RND	0.1733	0.1652 ± 0.0035	0.1651 ± 0.0034	7.6300 ± 0.8722
	FAS	0.1802	**0.1791 ± 0.0006**	**0.1790 ± 0.0006**	**5.7200 ± 0.5333**
	DFS	0.1795	0.1768 ± 0.0014	0.1767 ± 0.0014	5.9700 ± 0.4133
	BFT	**0.1817**	0.1786 ± 0.0022	0.1785 ± 0.0022	7.2100 ± 1.6532
Msweb	RND	0.1162	0.1116 ± 0.0024	0.1115 ± 0.0023	**3.4700 ± 2.0124**
	FAS	0.1227	**0.1219 ± 0.0004**	**0.1203 ± 0.0001**	24.1700 ± 3.4291
	DFS	0.1152	0.1127 ± 0.0009	0.1110 ± 0.0010	29.9000 ± 4.9615
	BFT	**0.1235**	0.1179 ± 0.0026	0.1176 ± 0.0026	15.8700 ± 59.9383
Book	RND	0.1168	0.1143 ± 0.0009	0.1143 ± 0.0009	1.8200 ± 0.3861
	FAS	**0.1220**	**0.1213 ± 0.0003**	**0.1213 ± 0.0003**	2.0000 ± 0.0000
	DFS	0.1201	0.1185 ± 0.0006	0.1185 ± 0.0006	1.7700 ± 0.4230
	BFT	0.1213	0.1202 ± 0.0005	0.1202 ± 0.0005	**1.7400 ± 0.4632**
Tmovie	RND	0.2883	0.2852 ± 0.0014	0.2851 ± 0.0014	2.0000 ± 0.0000
	FAS	**0.3034**	**0.3018 ± 0.0007**	**0.3018 ± 0.0007**	2.0000 ± 0.0000
	DFS	0.2955	0.2898 ± 0.0022	0.2898 ± 0.0021	2.0000 ± 0.0000
	BFT	0.2978	0.2950 ± 0.0011	0.2948 ± 0.0011	**2.0000 ± 0.0000**
Cwebkb	RND	0.1121	0.1090 ± 0.0014	0.1090 ± 0.0013	2.0000 ± 0.0000
	FAS	**0.1229**	**0.1219 ± 0.0005**	**0.1219 ± 0.0005**	2.0000 ± 0.0000
	DFS	0.1212	0.1196 ± 0.0007	0.1195 ± 0.0007	2.0000 ± 0.0000
	BFT	0.1191	0.1175 ± 0.0010	0.1174 ± 0.0010	**2.0000 ± 0.0000**
Cr52	RND	0.1449	0.1346 ± 0.0043	0.1345 ± 0.0042	2.0000 ± 0.0000
	FAS	**0.1592**	**0.1571 ± 0.0012**	**0.1571 ± 0.0012**	2.0000 ± 0.0000
	DFS	0.1574	0.1557 ± 0.0009	0.1556 ± 0.0009	2.0000 ± 0.0000
	BFT	0.1535	0.1489 ± 0.0026	0.1487 ± 0.0026	**2.0000 ± 0.0000**
C20ng	RND	0.0701	0.0680 ± 0.0010	0.0680 ± 0.0010	1.0000 ± 0.0000
	FAS	**0.0769**	**0.0764 ± 0.0003**	**0.0764 ± 0.0003**	1.0000 ± 0.0000
	DFS	0.0752	0.0742 ± 0.0005	0.0742 ± 0.0005	1.0000 ± 0.0000
	BFT	0.0747	0.0735 ± 0.0005	0.0735 ± 0.0005	**1.0000 ± 0.0000**
Bbc	RND	0.0680	0.0651 ± 0.0009	0.0651 ± 0.0009	2.2600 ± 0.4845
	FAS	**0.0768**	**0.0761 ± 0.0003**	**0.0761 ± 0.0003**	2.0000 ± 0.0000
	DFS	0.0733	0.0715 ± 0.0007	0.0715 ± 0.0007	**2.0000 ± 0.0000**
	BFT	0.0730	0.0718 ± 0.0006	0.0718 ± 0.0005	2.3400 ± 0.5360
Ad	RND	0.6909	0.6852 ± 0.0024	0.6851 ± 0.0023	2.2900 ± 0.8201
	FAS	0.6996	0.6949 ± 0.0018	0.6948 ± 0.0017	2.0000 ± 0.0000
	DFS	**0.6998**	**0.6958 ± 0.0019**	**0.6958 ± 0.0019**	**2.0000 ± 0.0000**
	BFT	0.6978	0.6915 ± 0.0025	0.6914 ± 0.0024	2.7300 ± 0.9413

Table E.1: *Best score obtained, Average best score obtained, Average initial score generated, Average number of iterations (Avg. It.) using sequential selection (best values in bold)*

Dataset	Approach	Best Score	Avg. Best Score	Avg. Initial Score	Avg. It.
Nltcs	RND	0.3410	0.3354 ± 0.0023	0.3323 ± 0.0027	7.2000 ± 3.0285
	FAS	**0.3415**	0.3401 ± 0.0010	**0.3379 ± 0.0010**	7.5500 ± 1.9611
	DFS	0.3414	**0.3409 ± 0.0003**	0.3373 ± 0.0008	**7.0900 ± 1.0160**
	BFT	0.3402	0.3353 ± 0.0027	0.3324 ± 0.0029	7.6900 ± 3.1741
Msnbc	RND	0.0574	0.0519 ± 0.0029	0.0498 ± 0.0031	97.7700 ± 189.4325
	FAS	0.0564	**0.0563 ± 0.0001**	0.0540 ± 0.0000	7.3000 ± 1.1934
	DFS	0.0553	0.0548 ± 0.0004	**0.0544 ± 0.0005**	**5.7700 ± 0.4230**
	BFT	**0.0581**	0.0525 ± 0.0030	0.0514 ± 0.0029	279.1600 ± 245.3886
Kdd	RND	0.1381	0.1360 ± 0.0011	0.1354 ± 0.0011	**8.5800 ± 2.7931**
	FAS	0.1401	**0.1395 ± 0.0003**	**0.1389 ± 0.0002**	36.2400 ± 106.9497
	DFS	0.1388	0.1381 ± 0.0003	0.1371 ± 0.0002	30.8400 ± 3.7732
	BFT	**0.1405**	0.1380 ± 0.0013	0.1373 ± 0.0014	63.4600 ± 154.2708
Plants	RND	0.5619	0.5575 ± 0.0019	0.5561 ± 0.0019	19.9900 ± 43.5019
	FAS	**0.5654**	**0.5639 ± 0.0006**	**0.5615 ± 0.0007**	**14.3500 ± 3.2979**
	DFS	0.5652	0.5627 ± 0.0009	0.5574 ± 0.0009	38.3400 ± 3.4089
	BFT	0.5644	0.5600 ± 0.0017	0.5584 ± 0.0018	15.8900 ± 35.8464
Baudio	RND	0.1563	0.1541 ± 0.0010	0.1537 ± 0.0009	24.7500 ± 20.8265
	FAS	0.1596	0.1576 ± 0.0009	0.1566 ± 0.0005	34.0700 ± 10.3468
	DFS	**0.1601**	**0.1594 ± 0.0004**	**0.1575 ± 0.0003**	37.2200 ± 4.2819
	BFT	0.1592	0.1573 ± 0.0008	0.1569 ± 0.0007	**18.8400 ± 18.7610**
Bnetflix	RND	0.1029	0.1010 ± 0.0008	0.1006 ± 0.0007	**24.6600 ± 33.4710**
	FAS	0.1044	0.1034 ± 0.0004	0.1021 ± 0.0004	26.1400 ± 3.9338
	DFS	**0.1059**	**0.1052 ± 0.0003**	**0.1038 ± 0.0004**	35.6100 ± 5.8429
	BFT	0.1018	0.1005 ± 0.0006	0.1003 ± 0.0006	27.0500 ± 35.2708
Jester	RND	0.1422	0.1396 ± 0.0011	0.1392 ± 0.0011	26.2400 ± 18.1036
	FAS	**0.1466**	**0.1458 ± 0.0004**	**0.1446 ± 0.0003**	**24.8800 ± 5.3149**
	DFS	0.1460	0.1449 ± 0.0007	0.1436 ± 0.0005	41.5600 ± 4.1178
	BFT	0.1421	0.1402 ± 0.0008	0.1398 ± 0.0008	25.2400 ± 31.0640
Accidents	RND	0.3741	0.3645 ± 0.0043	0.3635 ± 0.0042	32.8000 ± 98.6892
	FAS	0.3789	0.3749 ± 0.0016	0.3708 ± 0.0014	**30.3300 ± 4.4200**
	DFS	**0.3810**	**0.3790 ± 0.0009**	**0.3735 ± 0.0005**	45.8200 ± 5.1628
	BFT	0.3712	0.3647 ± 0.0035	0.3640 ± 0.0034	39.8900 ± 112.3748
Tretail	RND	0.0434	0.0421 ± 0.0008	0.0420 ± 0.0008	**3.8000 ± 1.6817**
	FAS	0.0442	**0.0440 ± 0.0001**	0.0430 ± 0.0001	17.7200 ± 2.9749
	DFS	0.0437	0.0434 ± 0.0002	0.0424 ± 0.0002	33.2900 ± 3.3403
	BFT	**0.0443**	0.0436 ± 0.0005	**0.0436 ± 0.0005**	4.1600 ± 1.9474
Pumsb_star	RND	0.6656	0.6539 ± 0.0057	0.6526 ± 0.0058	**14.8300 ± 20.1505**
	FAS	0.6734	0.6682 ± 0.0021	0.6662 ± 0.0019	26.9200 ± 6.2161
	DFS	**0.6756**	**0.6728 ± 0.0020**	**0.6703 ± 0.0021**	27.0000 ± 4.1463
	BFT	0.6596	0.6510 ± 0.0043	0.6502 ± 0.0042	27.4700 ± 32.2495
Dna	RND	0.1956	0.1941 ± 0.0008	0.1940 ± 0.0008	**262.5900 ± 248.3524**
	FAS	0.1979	0.1955 ± 0.0005	0.1953 ± 0.0002	482.9100 ± 84.1534
	DFS	**0.2008**	**0.1998 ± 0.0007**	**0.1990 ± 0.0003**	327.9600 ± 216.2747
	BFT	0.1974	0.1962 ± 0.0006	0.1962 ± 0.0006	311.7200 ± 241.7114
Kosarek	RND	0.1756	0.1669 ± 0.0034	0.1667 ± 0.0033	42.9700 ± 13.3941
	FAS	0.1727	0.1709 ± 0.0006	0.1708 ± 0.0005	**29.8700 ± 3.7407**
	DFS	0.1677	0.1666 ± 0.0006	0.1666 ± 0.0006	32.7500 ± 2.2174
	BFT	**0.1818**	**0.1780 ± 0.0023**	**0.1778 ± 0.0023**	47.5900 ± 32.3376
Msweb	RND	0.1171	0.1113 ± 0.0022	0.1111 ± 0.0022	**8.7300 ± 49.6551**
	FAS	0.1194	**0.1186 ± 0.0003**	**0.1170 ± 0.0001**	28.9000 ± 3.1766
	DFS	0.1129	0.1115 ± 0.0006	0.1098 ± 0.0007	40.1600 ± 3.8944
	BFT	**0.1204**	0.1170 ± 0.0019	0.1167 ± 0.0019	8.8300 ± 29.7685
Book	RND	0.1208	0.1189 ± 0.0008	0.1189 ± 0.0008	3.7400 ± 0.7604
	FAS	**0.1263**	**0.1258 ± 0.0002**	**0.1257 ± 0.0002**	**3.0000 ± 0.0000**
	DFS	0.1193	0.1183 ± 0.0004	0.1183 ± 0.0004	3.1600 ± 0.3685
	BFT	0.1260	0.1250 ± 0.0005	0.1249 ± 0.0005	3.6900 ± 0.6620
Tmovie	RND	0.3149	0.3116 ± 0.0017	0.3116 ± 0.0017	2.8700 ± 0.3667
	FAS	**0.3265**	**0.3247 ± 0.0010**	**0.3246 ± 0.0010**	**2.0000 ± 0.0000**
	DFS	0.3096	0.3063 ± 0.0017	0.3062 ± 0.0016	2.9300 ± 0.2564
	BFT	0.3263	0.3230 ± 0.0013	0.3230 ± 0.0012	2.0600 ± 0.2387
Cwebkb	RND	0.1265	0.1240 ± 0.0011	0.1239 ± 0.0011	2.2100 ± 0.4560
	FAS	0.1336	**0.1331 ± 0.0002**	**0.1331 ± 0.0002**	2.0000 ± 0.0000
	DFS	0.1219	0.1200 ± 0.0009	0.1200 ± 0.0009	**2.0000 ± 0.0000**
	BFT	**0.1338**	0.1322 ± 0.0007	0.1322 ± 0.0007	2.0800 ± 0.2727
Cr52	RND	0.1699	0.1606 ± 0.0033	0.1605 ± 0.0032	2.0300 ± 0.1714
	FAS	**0.1790**	**0.1780 ± 0.0003**	**0.1780 ± 0.0003**	**1.7100 ± 0.4560**
	DFS	0.1492	0.1475 ± 0.0010	0.1475 ± 0.0010	2.0000 ± 0.0000
	BFT	0.1790	0.1748 ± 0.0021	0.1747 ± 0.0021	2.1200 ± 0.4090
C20ng	RND	0.0888	0.0859 ± 0.0010	0.0859 ± 0.0010	2.0000 ± 0.0000
	FAS	0.0941	**0.0938 ± 0.0002**	**0.0938 ± 0.0001**	**1.5100 ± 0.5024**
	DFS	0.0853	0.0842 ± 0.0005	0.0842 ± 0.0005	1.9100 ± 0.2876
	BFT	**0.0949**	0.0937 ± 0.0006	0.0937 ± 0.0006	1.9900 ± 0.1000
Bbc	RND	0.0751	0.0731 ± 0.0008	0.0731 ± 0.0008	2.9100 ± 0.3208
	FAS	**0.0802**	**0.0799 ± 0.0001**	**0.0798 ± 0.0001**	2.0000 ± 0.0000
	DFS	0.0752	0.0734 ± 0.0007	0.0734 ± 0.0006	**2.0000 ± 0.0000**
	BFT	0.0770	0.0760 ± 0.0005	0.0760 ± 0.0005	2.3600 ± 0.4824
Ad	RND	0.7108	0.7066 ± 0.0015	0.7065 ± 0.0015	2.1100 ± 0.8152
	FAS	0.7137	0.7107 ± 0.0017	0.7106 ± 0.0016	2.0000 ± 0.0000
	DFS	**0.7176**	**0.7135 ± 0.0017**	**0.7134 ± 0.0017**	**2.0000 ± 0.0000**
	BFT	0.7146	0.7097 ± 0.0019	0.7095 ± 0.0018	2.5800 ± 0.7272

Table E.2: *Best score obtained, Average best score obtained, Average initial score generated, Average number of iterations (Avg. It.) using greedy selection (best values in bold)*

Dataset	Approach	Best Score	Avg. Best Score	Avg. Initial Score	Avg. It.
Nltcs	RND	0.3450	0.3439 ± 0.0007	0.3431 ± 0.0008	**7.5500 ± 3.5912**
	FAS	0.3455	0.3452 ± 0.0003	0.3436 ± 0.0000	12.9800 ± 1.8962
	DFS	0.3459	**0.3459 ± 0.0000**	**0.3443 ± 0.0002**	19.0000 ± 1.5505
	BFT	**0.3459**	0.3441 ± 0.0008	0.3431 ± 0.0010	8.9800 ± 3.8323
Msnbc	RND	**0.0823**	0.0776 ± 0.0031	0.0757 ± 0.0033	224.3500 ± 240.6807
	FAS	0.0771	0.0771 ± 0.0000	**0.0771 ± 0.0000**	500.0000 ± 0.0000
	DFS	0.0794	**0.0785 ± 0.0008**	0.0761 ± 0.0013	**18.9900 ± 5.4763**
	BFT	0.0817	0.0760 ± 0.0038	0.0739 ± 0.0040	291.0100 ± 241.8549
Kdd	RND	0.1332	0.1269 ± 0.0026	0.1260 ± 0.0028	**11.3100 ± 49.4278**
	FAS	0.1306	**0.1285 ± 0.0013**	0.1232 ± 0.0005	17.7200 ± 3.0849
	DFS	0.1280	0.1264 ± 0.0009	0.1181 ± 0.0002	37.5500 ± 4.4752
	BFT	**0.1336**	0.1281 ± 0.0028	**0.1269 ± 0.0027**	11.5900 ± 49.4223
Plants	RND	0.5368	0.5252 ± 0.0050	0.5228 ± 0.0051	71.5500 ± 166.4764
	FAS	0.5268	0.5213 ± 0.0021	0.5210 ± 0.0019	495.1300 ± 48.7000
	DFS	**0.5475**	**0.5456 ± 0.0008**	**0.5351 ± 0.0018**	**46.4800 ± 3.8310**
	BFT	0.5368	0.5279 ± 0.0043	0.5258 ± 0.0040	131.5500 ± 213.8157
Baudio	RND	0.1373	0.1300 ± 0.0028	0.1292 ± 0.0028	105.8100 ± 198.1070
	FAS	0.1402	0.1379 ± 0.0014	0.1311 ± 0.0003	46.6500 ± 93.0530
	DFS	0.1326	0.1301 ± 0.0013	0.1220 ± 0.0014	**42.6000 ± 3.8925**
	BFT	**0.1429**	**0.1382 ± 0.0022**	**0.1378 ± 0.0021**	69.8900 ± 167.1093
Bnetflix	RND	0.0928	0.0871 ± 0.0024	0.0864 ± 0.0024	61.4000 ± 154.9891
	FAS	0.0924	**0.0910 ± 0.0006**	0.0872 ± 0.0001	**28.1800 ± 3.1152**
	DFS	0.0886	0.0863 ± 0.0011	0.0808 ± 0.0012	47.5000 ± 5.4170
	BFT	**0.0947**	0.0899 ± 0.0019	**0.0895 ± 0.0019**	65.4300 ± 161.2970
Jester	RND	0.1239	0.1149 ± 0.0034	0.1144 ± 0.0033	**308.4100 ± 240.8296**
	FAS	**0.1263**	**0.1214 ± 0.0017**	**0.1201 ± 0.0004**	466.4200 ± 123.0152
	DFS	0.1187	0.1109 ± 0.0041	0.1078 ± 0.0010	333.5900 ± 218.2524
	BFT	0.1247	0.1193 ± 0.0027	0.1190 ± 0.0027	381.9000 ± 211.2247
Accidents	RND	0.3164	0.3005 ± 0.0080	0.2994 ± 0.0079	144.5800 ± 222.7731
	FAS	**0.3333**	**0.3255 ± 0.0031**	**0.3204 ± 0.0007**	**18.0100 ± 3.7592**
	DFS	0.3257	0.3214 ± 0.0023	0.3044 ± 0.0013	196.0800 ± 219.2677
	BFT	0.3182	0.3062 ± 0.0058	0.3055 ± 0.0056	55.3400 ± 148.9850
Tretail	RND	0.0437	0.0408 ± 0.0019	0.0408 ± 0.0019	28.3600 ± 108.7575
	FAS	0.0442	**0.0439 ± 0.0001**	**0.0436 ± 0.0000**	**16.6800 ± 2.6888**
	DFS	0.0423	0.0398 ± 0.0012	0.0365 ± 0.0001	33.2600 ± 4.4803
	BFT	**0.0444**	0.0434 ± 0.0008	0.0434 ± 0.0008	19.1700 ± 85.0037
Pumsb_star	RND	0.6197	0.6059 ± 0.0060	0.6048 ± 0.0059	169.0800 ± 233.4214
	FAS	0.6396	**0.6352 ± 0.0026**	0.6301 ± 0.0015	**46.4400 ± 125.0822**
	DFS	**0.6416**	0.6348 ± 0.0026	**0.6337 ± 0.0022**	399.3100 ± 196.2845
	BFT	0.6151	0.6018 ± 0.0057	0.6009 ± 0.0054	124.6800 ± 211.9875
Dna	RND	0.1903	0.1885 ± 0.0008	0.1883 ± 0.0008	301.5400 ± 244.2900
	FAS	**0.1962**	**0.1946 ± 0.0004**	**0.1945 ± 0.0001**	490.9900 ± 63.3878
	DFS	0.1959	0.1935 ± 0.0009	0.1924 ± 0.0007	477.1700 ± 100.0209
	BFT	0.1921	0.1905 ± 0.0006	0.1904 ± 0.0006	**276.3900 ± 248.4581**
Kosarek	RND	0.1690	0.1580 ± 0.0049	0.1577 ± 0.0048	**360.1200 ± 219.9985**
	FAS	0.1740	0.1704 ± 0.0014	0.1700 ± 0.0010	419.5100 ± 127.5524
	DFS	0.1577	0.1536 ± 0.0017	0.1530 ± 0.0015	404.7300 ± 74.3770
	BFT	**0.1774**	**0.1712 ± 0.0031**	**0.1710 ± 0.0030**	378.0800 ± 212.2504
Msweb	RND	0.1124	0.1068 ± 0.0028	0.1066 ± 0.0028	85.5700 ± 177.2279
	FAS	0.1194	**0.1183 ± 0.0004**	**0.1166 ± 0.0000**	**24.0400 ± 4.2945**
	DFS	0.1074	0.1010 ± 0.0029	0.0939 ± 0.0024	63.6700 ± 74.4034
	BFT	**0.1204**	0.1140 ± 0.0028	0.1134 ± 0.0027	78.6200 ± 164.7486
Book	RND	0.1182	0.1146 ± 0.0011	0.1145 ± 0.0011	8.7300 ± 1.8848
	FAS	0.1217	0.1211 ± 0.0003	0.1211 ± 0.0002	7.0100 ± 0.5024
	DFS	0.1124	0.1108 ± 0.0007	0.1108 ± 0.0007	7.9000 ± 0.3624
	BFT	**0.1236**	**0.1218 ± 0.0008**	**0.1216 ± 0.0007**	**6.7800 ± 2.4229**
Tmovie	RND	0.2854	0.2764 ± 0.0038	0.2762 ± 0.0037	12.8300 ± 2.8641
	FAS	0.2893	0.2877 ± 0.0006	0.2876 ± 0.0006	10.2200 ± 0.4623
	DFS	0.2580	0.2502 ± 0.0040	0.2502 ± 0.0040	**9.5500 ± 0.5000**
	BFT	**0.3067**	**0.2982 ± 0.0038**	**0.2979 ± 0.0037**	13.3400 ± 2.7897
Cwebkb	RND	0.1204	0.1164 ± 0.0017	0.1163 ± 0.0016	4.1100 ± 0.8027
	FAS	0.1249	0.1247 ± 0.0001	0.1246 ± 0.0001	3.0000 ± 0.0000
	DFS	0.1120	0.1102 ± 0.0007	0.1102 ± 0.0007	**3.0000 ± 0.0000**
	BFT	**0.1289**	**0.1265 ± 0.0012**	**0.1264 ± 0.0012**	4.1200 ± 0.9876
Cr52	RND	0.1581	0.1445 ± 0.0054	0.1443 ± 0.0052	4.2900 ± 0.5183
	FAS	0.1635	**0.1630 ± 0.0003**	**0.1629 ± 0.0003**	**3.0000 ± 0.0000**
	DFS	0.1408	0.1309 ± 0.0028	0.1308 ± 0.0026	3.0100 ± 0.1000
	BFT	**0.1691**	0.1622 ± 0.0034	0.1620 ± 0.0034	3.9600 ± 0.4907
C20ng	RND	0.0772	0.0717 ± 0.0019	0.0717 ± 0.0018	3.0600 ± 0.2387
	FAS	0.0825	0.0823 ± 0.0001	0.0823 ± 0.0001	2.7500 ± 0.4352
	DFS	0.0610	0.0593 ± 0.0008	0.0593 ± 0.0008	**2.0000 ± 0.0000**
	BFT	**0.0868**	**0.0832 ± 0.0018**	**0.0831 ± 0.0018**	3.2000 ± 0.6963
Bbc	RND	0.0731	0.0709 ± 0.0009	0.0708 ± 0.0009	4.0000 ± 0.5125
	FAS	**0.0792**	**0.0790 ± 0.0001**	**0.0790 ± 0.0001**	2.9800 ± 0.1407
	DFS	0.0736	0.0716 ± 0.0010	0.0715 ± 0.0010	**2.9600 ± 0.1969**
	BFT	0.0751	0.0740 ± 0.0006	0.0740 ± 0.0006	4.0500 ± 0.5198
Ad	RND	0.6978	0.6934 ± 0.0023	0.6933 ± 0.0023	2.1900 ± 0.8127
	FAS	**0.7049**	**0.7015 ± 0.0017**	**0.7014 ± 0.0016**	2.0000 ± 0.0000
	DFS	0.7035	0.6998 ± 0.0018	0.6998 ± 0.0018	**2.0000 ± 0.0000**
	BFT	0.7036	0.6993 ± 0.0019	0.6991 ± 0.0020	2.5300 ± 0.7447

Table E.3: *Best score obtained, Average best score obtained, Average initial score generated, Average number of iterations (Avg. It.) using independence selection (best values in bold)*

Appendix F

Empirical Results: ASOBS

Dataset	Approach	Best Score	Avg. Best Score	Avg. Initial Score	Avg. It.
Nltcs	RND	0.3458	0.3443 ± 0.0009	0.3435 ± 0.0009	4.7500 ± 2.5638
	FAS	0.3457	0.3451 ± 0.0004	0.3436 ± 0.0000	6.5100 ± 2.5839
	DFS	0.3459	**0.3459 ± 0.0000**	**0.3443 ± 0.0002**	9.4500 ± 1.1404
	BFT	**0.3462**	0.3444 ± 0.0007	0.3436 ± 0.0008	**4.0400 ± 1.7974**
Msnbc	RND	0.0826	0.0808 ± 0.0009	0.0798 ± 0.0011	6.7700 ± 3.0612
	FAS	0.0826	**0.0826 ± 0.0000**	**0.0817 ± 0.0000**	**5.0000 ± 0.0000**
	DFS	0.0804	0.0804 ± 0.0000	0.0788 ± 0.0003	17.9900 ± 0.6889
	BFT	**0.0832**	0.0815 ± 0.0010	0.0807 ± 0.0012	5.0800 ± 2.4931
Kdd	RND	**0.1504**	0.1447 ± 0.0016	0.1438 ± 0.0014	**3.7000 ± 1.8614**
	FAS	0.1467	**0.1459 ± 0.0005**	0.1442 ± 0.0002	15.7700 ± 2.5380
	DFS	0.1469	0.1449 ± 0.0006	0.1434 ± 0.0005	20.1200 ± 3.6328
	BFT	0.1473	0.1453 ± 0.0009	**0.1445 ± 0.0009**	5.0600 ± 2.2555
Plants	RND	**0.5839**	0.5761 ± 0.0026	0.5738 ± 0.0015	6.4600 ± 2.7355
	FAS	0.5796	0.5764 ± 0.0014	0.5730 ± 0.0004	18.8600 ± 2.9679
	DFS	0.5796	0.5760 ± 0.0011	0.5718 ± 0.0008	28.5800 ± 4.9076
	BFT	0.5824	**0.5777 ± 0.0016**	**0.5760 ± 0.0012**	**6.0200 ± 2.8989**
Baudio	RND	0.1640	0.1622 ± 0.0011	0.1611 ± 0.0012	**4.1800 ± 1.9867**
	FAS	**0.1654**	0.1632 ± 0.0012	0.1591 ± 0.0002	26.6900 ± 4.1407
	DFS	0.1638	0.1614 ± 0.0010	0.1586 ± 0.0005	26.9600 ± 4.4967
	BFT	0.1654	**0.1633 ± 0.0010**	**0.1621 ± 0.0012**	4.1900 ± 1.9576
Bnetflix	RND	0.1103	0.1079 ± 0.0009	0.1071 ± 0.0011	4.0900 ± 2.0256
	FAS	**0.1112**	**0.1096 ± 0.0006**	0.1065 ± 0.0004	27.7100 ± 3.6578
	DFS	0.1093	0.1080 ± 0.0006	0.1057 ± 0.0005	24.4000 ± 4.4290
	BFT	0.1100	0.1078 ± 0.0009	**0.1072 ± 0.0009**	**3.0700 ± 1.6034**
Jester	RND	0.1490	0.1466 ± 0.0010	0.1457 ± 0.0011	**4.1100 ± 1.9222**
	FAS	0.1486	**0.1475 ± 0.0005**	0.1450 ± 0.0003	24.8700 ± 3.8446
	DFS	0.1478	0.1461 ± 0.0007	0.1447 ± 0.0006	24.2700 ± 3.4256
	BFT	**0.1492**	0.1471 ± 0.0009	**0.1461 ± 0.0011**	5.0500 ± 2.1667
Accidents	RND	0.3869	0.3747 ± 0.0050	0.3726 ± 0.0048	3.1200 ± 1.8764
	FAS	0.3916	0.3788 ± 0.0036	0.3714 ± 0.0019	25.6600 ± 4.0806
	DFS	0.3966	**0.3833 ± 0.0062**	0.3702 ± 0.0016	34.1900 ± 3.5553
	BFT	**0.3984**	0.3797 ± 0.0052	**0.3781 ± 0.0051**	**2.5700 ± 1.6712**
Tretail	RND	0.0445	0.0441 ± 0.0002	0.0441 ± 0.0002	**1.1000 ± 1.0493**
	FAS	0.0446	0.0444 ± 0.0001	0.0440 ± 0.0000	16.9000 ± 2.5485
	DFS	0.0446	0.0444 ± 0.0001	0.0425 ± 0.0001	31.0800 ± 3.3233
	BFT	**0.0448**	**0.0445 ± 0.0002**	**0.0444 ± 0.0002**	1.8200 ± 1.3056
Pumsb_star	RND	0.7362	0.7207 ± 0.0059	0.7197 ± 0.0057	**1.5700 ± 1.2812**
	FAS	**0.7434**	**0.7309 ± 0.0059**	0.7124 ± 0.0050	34.3220 ± 4.2647
	DFS	0.7308	0.7173 ± 0.0051	0.6981 ± 0.0040	43.5696 ± 4.1685
	BFT	0.7398	0.7257 ± 0.0058	**0.7246 ± 0.0059**	1.6300 ± 1.1340
Dna	RND	0.1992	0.1980 ± 0.0005	0.1979 ± 0.0005	1.9500 ± 1.4311
	FAS	0.2000	0.1995 ± 0.0002	0.1970 ± 0.0002	54.8889 ± 5.3333
	DFS	**0.2011**	**0.2008 ± 0.0001**	**0.1992 ± 0.0002**	62.0217 ± 6.6650
	BFT	0.1991	0.1986 ± 0.0003	0.1986 ± 0.0003	**1.1600 ± 1.0222**
Kosarek	RND	0.1826	0.1799 ± 0.0015	0.1795 ± 0.0015	**1.6900 ± 1.2689**
	FAS	0.1842	0.1831 ± 0.0006	0.1802 ± 0.0006	24.5700 ± 2.9857
	DFS	0.1839	0.1831 ± 0.0005	0.1804 ± 0.0010	34.4750 ± 4.5006
	BFT	**0.1864**	**0.1835 ± 0.0012**	**0.1829 ± 0.0013**	2.6100 ± 1.7747
Msweb	RND	0.1248	0.1216 ± 0.0012	0.1215 ± 0.0012	**0.8090 ± 0.8101**
	FAS	0.1244	0.1234 ± 0.0007	0.1209 ± 0.0001	27.4545 ± 4.2980
	DFS	0.1238	0.1230 ± 0.0007	0.1194 ± 0.0006	31.2857 ± 3.4503
	BFT	**0.1251**	**0.1236 ± 0.0010**	**0.1234 ± 0.0009**	2.2143 ± 1.4573
Book	RND	0.1244	0.1237 ± 0.0004	0.1237 ± 0.0004	**0.8889 ± 1.0500**
	FAS	0.1247	0.1244 ± 0.0002	0.1235 ± 0.0002	21.3333 ± 1.5811
	DFS	0.1246	0.1244 ± 0.0003	0.1230 ± 0.0003	24.7500 ± 4.9917
	BFT	**0.1252**	**0.1247 ± 0.0003**	**0.1246 ± 0.0003**	2.1905 ± 1.2091
Tmovie	RND	0.3101	0.3080 ± 0.0011	0.3079 ± 0.0011	**0.3000 ± 0.5960**
	FAS	0.3102	**0.3093 ± 0.0007**	0.3054 ± 0.0012	13.5714 ± 1.1339
	DFS	0.3082	0.3072 ± 0.0015	0.3034 ± 0.0004	23.5000 ± 0.7071
	BFT	**0.3104**	0.3089 ± 0.0009	**0.3087 ± 0.0009**	1.5152 ± 1.3257
Cwebkb	RND	0.1260	0.1252 ± 0.0007	0.1252 ± 0.0007	**0.0000 ± 0.0000**
	FAS	**0.1260**	0.1257 ± 0.0005	0.1252 ± 0.0012	5.0000 ± 4.2426
	DFS	0.1258	0.1258 ± 0.0000	0.1254 ± 0.0005	5.0000 ± 4.2426
	BFT	0.1260	**0.1259 ± 0.0002**	**0.1259 ± 0.0002**	2.0000 ± 0.0000
Cr52	RND	0.1616	0.1607 ± 0.0010	0.1607 ± 0.0010	**0.2500 ± 0.5000**
	FAS	0.1615	0.1593 ± 0.0019	0.1442 ± 0.0137	6.6667 ± 2.5166
	DFS	0.1595	0.1582 ± 0.0009	0.1342 ± 0.0035	6.2857 ± 2.4113
	BFT	**0.1638**	**0.1620 ± 0.0014**	**0.1620 ± 0.0014**	0.3333 ± 0.5000
C20ng	RND	0.0778	0.0775 ± 0.0003	0.0775 ± 0.0003	**0.2500 ± 0.5000**
	FAS	**0.0781**	**0.0781 ± 0.0000**	0.0775 ± 0.0002	5.5000 ± 0.7071
	DFS	0.0778	0.0778 ± 0.0000	0.0771 ± 0.0000	6.0000 ± 0.0000
	BFT	0.0780	0.0777 ± 0.0002	**0.0777 ± 0.0002**	0.8333 ± 1.1690
Bbc	RND	0.0781	0.0781 ± 0.0000	0.0781 ± 0.0000	1.0000 ± 0.0000
	FAS	0.0780	0.0780 ± 0.0000	0.0778 ± 0.0000	2.0000 ± 0.0000
	DFS	0.0778	0.0778 ± 0.0000	0.0775 ± 0.0000	2.0000 ± 0.0000
	BFT	**0.0787**	**0.0787 ± 0.0000**	**0.0786 ± 0.0000**	**0.0000 ± 0.0000**
Ad	RND	0.7518	0.7518 ± 0.0000	0.7518 ± 0.0000	0.0000 ± 0.0000
	FAS	0.7272	0.7272 ± 0.0000	0.7272 ± 0.0000	0.0000 ± 0.0000
	DFS	0.7196	0.7196 ± 0.0000	0.7196 ± 0.0000	0.0000 ± 0.0000
	BFT	**0.7544**	**0.7531 ± 0.0018**	**0.7531 ± 0.0018**	**0.0000 ± 0.0000**

Table F.1: *Best score obtained, Average best score obtained, Average initial score generated, Average number of iterations (Avg. It.) using sequential selection (best values in bold)*

Dataset	Approach	Best Score	Avg. Best Score	Avg. Initial Score	Avg. It.
Nltcs	RND	0.3584	0.3414 ± 0.0051	0.3364 ± 0.0045	3.2100 ± 1.8494
	FAS	0.3489	**0.3458 ± 0.0031**	**0.3382 ± 0.0010**	3.5500 ± 0.6571
	DFS	0.3416	0.3411 ± 0.0003	0.3377 ± 0.0007	**2.4100 ± 0.8052**
	BFT	**0.3651**	0.3434 ± 0.0065	0.3362 ± 0.0038	3.2300 ± 2.0244
Msnbc	RND	0.0611	0.0564 ± 0.0021	0.0536 ± 0.0025	3.4600 ± 1.9250
	FAS	0.0576	0.0573 ± 0.0003	**0.0569 ± 0.0000**	**2.3500 ± 1.1044**
	DFS	0.0590	**0.0590 ± 0.0001**	0.0547 ± 0.0004	6.1400 ± 0.9951
	BFT	**0.0619**	0.0580 ± 0.0023	0.0555 ± 0.0025	3.2900 ± 1.7425
Kdd	RND	0.1462	0.1419 ± 0.0018	0.1407 ± 0.0018	**3.3200 ± 1.3845**
	FAS	**0.1479**	**0.1435 ± 0.0010**	0.1407 ± 0.0007	14.8900 ± 2.5342
	DFS	0.1452	0.1423 ± 0.0013	0.1384 ± 0.0006	16.6700 ± 2.5978
	BFT	0.1471	0.1424 ± 0.0018	**0.1409 ± 0.0016**	4.6600 ± 2.2483
Plants	RND	**0.5893**	0.5743 ± 0.0059	0.5681 ± 0.0049	**4.8900 ± 2.3002**
	FAS	0.5831	0.5742 ± 0.0038	0.5632 ± 0.0008	15.0300 ± 2.6684
	DFS	0.5759	0.5715 ± 0.0019	0.5607 ± 0.0011	23.8100 ± 4.6006
	BFT	0.5888	**0.5764 ± 0.0049**	**0.5708 ± 0.0047**	4.9900 ± 2.4308
Baudio	RND	0.1662	0.1632 ± 0.0012	0.1615 ± 0.0014	5.0700 ± 2.2930
	FAS	0.1657	0.1636 ± 0.0011	0.1584 ± 0.0005	22.3700 ± 4.3939
	DFS	0.1665	0.1642 ± 0.0008	0.1598 ± 0.0005	21.6800 ± 3.9616
	BFT	**0.1674**	**0.1643 ± 0.0014**	**0.1628 ± 0.0014**	**4.6300 ± 2.5649**
Bnetflix	RND	0.1101	0.1078 ± 0.0010	**0.1067 ± 0.0010**	**4.3800 ± 1.9216**
	FAS	**0.1105**	0.1081 ± 0.0009	0.1038 ± 0.0004	25.6200 ± 3.8919
	DFS	0.1099	**0.1085 ± 0.0007**	0.1058 ± 0.0008	19.3900 ± 3.7088
	BFT	0.1098	0.1074 ± 0.0010	0.1062 ± 0.0011	4.4000 ± 1.9540
Jester	RND	0.1498	0.1474 ± 0.0012	0.1461 ± 0.0013	4.7200 ± 2.2745
	FAS	**0.1511**	**0.1494 ± 0.0008**	0.1462 ± 0.0006	20.0700 ± 3.8488
	DFS	0.1503	0.1485 ± 0.0005	0.1453 ± 0.0006	27.1500 ± 4.2268
	BFT	0.1509	0.1479 ± 0.0012	**0.1464 ± 0.0013**	**4.6900 ± 1.9730**
Accidents	RND	0.4005	0.3833 ± 0.0057	0.3812 ± 0.0049	3.4000 ± 1.7408
	FAS	**0.4044**	**0.3891 ± 0.0051**	0.3768 ± 0.0025	25.2400 ± 3.9392
	DFS	0.3854	0.3840 ± 0.0006	0.3775 ± 0.0006	28.1100 ± 2.9675
	BFT	0.4032	0.3877 ± 0.0053	**0.3860 ± 0.0054**	**3.0000 ± 1.9069**
Tretail	RND	0.0444	0.0436 ± 0.0006	0.0435 ± 0.0006	**0.9900 ± 1.0588**
	FAS	0.0446	0.0443 ± 0.0001	0.0432 ± 0.0001	16.1800 ± 2.4919
	DFS	0.0446	**0.0445 ± 0.0001**	0.0435 ± 0.0001	28.5200 ± 3.7483
	BFT	**0.0448**	0.0443 ± 0.0002	**0.0443 ± 0.0002**	1.4100 ± 1.1379
Pumsb_star	RND	0.7200	0.7048 ± 0.0060	0.7037 ± 0.0057	**1.2600 ± 1.0975**
	FAS	0.7192	**0.7089 ± 0.0051**	0.6917 ± 0.0039	27.9077 ± 3.8190
	DFS	0.7167	0.7049 ± 0.0052	0.6812 ± 0.0034	33.3800 ± 4.1652
	BFT	**0.7208**	0.7071 ± 0.0068	**0.7061 ± 0.0067**	1.2900 ± 1.0852
Dna	RND	0.1989	0.1977 ± 0.0005	0.1976 ± 0.0005	1.7500 ± 1.2092
	FAS	0.1996	0.1990 ± 0.0003	0.1965 ± 0.0002	59.8571 ± 7.3991
	DFS	**0.2017**	**0.2014 ± 0.0001**	**0.1993 ± 0.0002**	59.9792 ± 5.9804
	BFT	0.1996	0.1989 ± 0.0004	0.1989 ± 0.0004	**1.1200 ± 1.0375**
Kosarek	RND	0.1813	0.1782 ± 0.0016	0.1778 ± 0.0015	**2.0900 ± 1.5641**
	FAS	0.1809	0.1794 ± 0.0006	0.1740 ± 0.0008	34.8846 ± 4.1504
	DFS	0.1793	0.1779 ± 0.0008	0.1722 ± 0.0008	46.8333 ± 5.8089
	BFT	**0.1843**	**0.1822 ± 0.0011**	**0.1818 ± 0.0011**	2.9200 ± 1.7273
Msweb	RND	0.1236	0.1204 ± 0.0012	0.1202 ± 0.0012	**1.0204 ± 0.9893**
	FAS	0.1224	0.1216 ± 0.0004	0.1177 ± 0.0001	27.7143 ± 2.6859
	DFS	0.1220	0.1218 ± 0.0004	0.1180 ± 0.0008	37.5000 ± 4.9497
	BFT	**0.1259**	**0.1224 ± 0.0013**	**0.1218 ± 0.0013**	2.4200 ± 1.4995
Book	RND	0.1291	0.1280 ± 0.0004	0.1280 ± 0.0004	1.4737 ± 1.2635
	FAS	0.1275	0.1274 ± 0.0001	0.1274 ± 0.0001	**0.0000 ± 0.0000**
	DFS	0.1278	0.1278 ± 0.0000	0.1259 ± 0.0001	26.5000 ± 0.7071
	BFT	**0.1296**	**0.1290 ± 0.0003**	**0.1288 ± 0.0004**	3.4000 ± 2.0111
Tmovie	RND	0.3376	0.3349 ± 0.0016	0.3344 ± 0.0016	**1.2308 ± 1.1422**
	FAS	0.3378	0.3359 ± 0.0012	0.3301 ± 0.0008	12.1176 ± 0.6002
	DFS	0.3336	0.3332 ± 0.0005	0.3289 ± 0.0010	27.5000 ± 3.5355
	BFT	**0.3389**	**0.3364 ± 0.0013**	**0.3349 ± 0.0013**	3.9524 ± 1.9121
Cwebkb	RND	0.1385	0.1382 ± 0.0003	0.1381 ± 0.0002	**0.4000 ± 0.5477**
	FAS	0.1377	0.1371 ± 0.0005	0.1361 ± 0.0005	5.0000 ± 0.0000
	DFS	0.1366	0.1364 ± 0.0003	0.1345 ± 0.0004	6.0000 ± 0.0000
	BFT	**0.1397**	**0.1397 ± 0.0001**	**0.1397 ± 0.0000**	1.0000 ± 1.4142
Cr52	RND	0.1851	0.1841 ± 0.0013	0.1841 ± 0.0013	**0.0000 ± 0.0000**
	FAS	0.1845	0.1837 ± 0.0005	0.1816 ± 0.0004	4.0000 ± 0.0000
	DFS	0.1802	0.1787 ± 0.0016	0.1767 ± 0.0014	4.5000 ± 0.5774
	BFT	**0.1866**	**0.1866 ± 0.0000**	**0.1860 ± 0.0000**	6.0000 ± 0.0000
C20ng	RND	0.0974	0.0970 ± 0.0004	0.0970 ± 0.0003	**1.2500 ± 0.5000**
	FAS	0.0964	0.0962 ± 0.0001	0.0958 ± 0.0001	2.6667 ± 0.4924
	DFS	0.0956	0.0951 ± 0.0005	0.0945 ± 0.0005	2.8333 ± 0.4082
	BFT	**0.0984**	**0.0980 ± 0.0003**	**0.0978 ± 0.0002**	3.2500 ± 1.6690
Bbc	RND	0.0844	0.0844 ± 0.0000	0.0844 ± 0.0000	0.0000 ± 0.0000
	FAS	0.0828	0.0828 ± 0.0000	0.0824 ± 0.0002	3.0000 ± 0.0000
	DFS	0.0827	0.0827 ± 0.0000	0.0823 ± 0.0000	3.0000 ± 0.0000
	BFT	**0.0846**	**0.0846 ± 0.0000**	**0.0846 ± 0.0000**	**0.0000 ± 0.0000**
Ad	RND	0.7649	0.7649 ± 0.0000	0.7649 ± 0.0000	**0.0000 ± 0.0000**
	FAS	0.7549	0.7549 ± 0.0000	0.7539 ± 0.0000	1.0000 ± 0.0000
	DFS	0.7381	0.7381 ± 0.0000	0.7353 ± 0.0000	2.0000 ± 0.0000
	BFT	**0.7750**	**0.7750 ± 0.0000**	**0.7747 ± 0.0000**	1.0000 ± 0.0000

Table F.2: *Best score obtained, Average best score obtained, Average initial score generated, Average number of iterations (Avg. It.) using greedy selection (best values in bold)*

Dataset	Approach	Best Score	Avg. Best Score	Avg. Initial Score	Avg. It.
Nltcs	RND	0.3470	0.3444 ± 0.0009	0.3435 ± 0.0008	4.5500 ± 2.5201
	FAS	0.3457	0.3451 ± 0.0003	0.3436 ± 0.0000	7.1300 ± 1.7959
	DFS	0.3459	**0.3459 ± 0.0000**	**0.3443 ± 0.0002**	9.5200 ± 1.1054
	BFT	**0.3491**	0.3444 ± 0.0010	0.3435 ± 0.0007	**4.2100 ± 2.2710**
Msnbc	RND	0.0819	0.0795 ± 0.0019	0.0775 ± 0.0028	4.3900 ± 2.3437
	FAS	0.0800	**0.0799 ± 0.0002**	**0.0791 ± 0.0000**	6.2100 ± 2.1048
	DFS	0.0801	0.0797 ± 0.0004	0.0775 ± 0.0008	7.1900 ± 1.4404
	BFT	**0.0823**	0.0792 ± 0.0017	0.0772 ± 0.0022	**3.9100 ± 2.3359**
Kdd	RND	0.1396	0.1360 ± 0.0018	**0.1353 ± 0.0018**	**1.7600 ± 1.3266**
	FAS	0.1367	0.1351 ± 0.0007	0.1304 ± 0.0002	9.7200 ± 1.9179
	DFS	0.1397	**0.1384 ± 0.0004**	0.1279 ± 0.0016	18.5100 ± 1.8504
	BFT	**0.1415**	0.1361 ± 0.0022	0.1347 ± 0.0022	2.7000 ± 1.6049
Plants	RND	0.5699	0.5556 ± 0.0057	0.5510 ± 0.0056	**2.7100 ± 1.5262**
	FAS	0.5596	0.5491 ± 0.0036	0.5409 ± 0.0019	10.7200 ± 2.8677
	DFS	**0.5813**	**0.5728 ± 0.0038**	**0.5548 ± 0.0021**	17.6800 ± 3.0114
	BFT	0.5712	0.5563 ± 0.0053	0.5508 ± 0.0052	3.2800 ± 1.8372
Baudio	RND	0.1480	0.1441 ± 0.0019	0.1436 ± 0.0019	1.8500 ± 1.3437
	FAS	0.1480	0.1466 ± 0.0007	0.1423 ± 0.0002	17.6000 ± 2.3826
	DFS	0.1501	**0.1471 ± 0.0010**	0.1368 ± 0.0018	27.2300 ± 3.9231
	BFT	**0.1504**	0.1461 ± 0.0018	**0.1457 ± 0.0018**	**1.5000 ± 1.2350**
Bnetflix	RND	0.1005	0.0977 ± 0.0013	0.0973 ± 0.0013	1.7800 ± 1.4040
	FAS	0.0994	0.0985 ± 0.0004	0.0959 ± 0.0002	14.5700 ± 2.2077
	DFS	0.1012	**0.1004 ± 0.0004**	0.0965 ± 0.0007	19.1000 ± 3.4068
	BFT	**0.1014**	0.0984 ± 0.0011	**0.0982 ± 0.0011**	**1.4400 ± 1.1399**
Jester	RND	0.1348	0.1312 ± 0.0018	0.1307 ± 0.0018	1.6000 ± 1.3633
	FAS	0.1348	0.1328 ± 0.0007	0.1296 ± 0.0001	10.0800 ± 2.0532
	DFS	**0.1364**	**0.1344 ± 0.0009**	0.1304 ± 0.0013	15.1900 ± 3.2805
	BFT	0.1359	0.1324 ± 0.0015	**0.1321 ± 0.0014**	**1.3900 ± 1.1712**
Accidents	RND	0.3544	0.3386 ± 0.0060	0.3371 ± 0.0060	1.2600 ± 1.1247
	FAS	**0.3573**	**0.3488 ± 0.0040**	**0.3410 ± 0.0038**	12.7800 ± 2.2137
	DFS	0.3506	0.3443 ± 0.0028	0.3317 ± 0.0011	23.2500 ± 2.7280
	BFT	0.3546	0.3414 ± 0.0064	0.3402 ± 0.0063	**0.7900 ± 0.9134**
Tretail	RND	0.0445	0.0440 ± 0.0002	0.0440 ± 0.0002	**1.1000 ± 1.0964**
	FAS	0.0445	0.0443 ± 0.0001	0.0440 ± 0.0001	14.1700 ± 2.2341
	DFS	0.0443	0.0442 ± 0.0001	0.0434 ± 0.0001	29.0100 ± 2.9627
	BFT	**0.0446**	**0.0443 ± 0.0002**	**0.0443 ± 0.0002**	1.7100 ± 1.2736
Pumsb_star	RND	0.6894	0.6718 ± 0.0071	0.6707 ± 0.0067	**0.7400 ± 0.8241**
	FAS	0.6826	0.6719 ± 0.0054	0.6527 ± 0.0024	14.3900 ± 2.5222
	DFS	**0.6896**	**0.6792 ± 0.0054**	0.6551 ± 0.0038	21.1100 ± 2.8706
	BFT	0.6887	0.6746 ± 0.0067	**0.6737 ± 0.0066**	0.9000 ± 0.9898
Dna	RND	0.1941	0.1928 ± 0.0007	0.1927 ± 0.0007	1.0600 ± 1.0330
	FAS	0.1966	0.1963 ± 0.0001	**0.1949 ± 0.0000**	42.6250 ± 3.6281
	DFS	**0.1972**	**0.1966 ± 0.0003**	0.1929 ± 0.0007	42.0370 ± 5.9127
	BFT	0.1955	0.1943 ± 0.0005	0.1943 ± 0.0005	**0.5000 ± 0.6742**
Kosarek	RND	0.1803	0.1757 ± 0.0017	0.1754 ± 0.0017	**1.0200 ± 0.9209**
	FAS	**0.1823**	**0.1802 ± 0.0011**	0.1752 ± 0.0015	27.8205 ± 3.5383
	DFS	0.1777	0.1760 ± 0.0010	0.1710 ± 0.0020	29.8000 ± 3.7859
	BFT	0.1802	0.1783 ± 0.0012	**0.1780 ± 0.0011**	1.4300 ± 1.2654
Msweb	RND	0.1214	0.1190 ± 0.0016	0.1189 ± 0.0016	**0.5577 ± 0.6690**
	FAS	0.1215	0.1212 ± 0.0003	0.1196 ± 0.0000	22.1250 ± 3.1820
	DFS	0.1217	**0.1214 ± 0.0003**	0.1152 ± 0.0014	29.3333 ± 4.0415
	BFT	**0.1229**	0.1206 ± 0.0014	**0.1203 ± 0.0014**	1.1622 ± 0.9512
Book	RND	0.1260	0.1252 ± 0.0004	0.1251 ± 0.0004	1.2667 ± 1.0328
	FAS	0.1259	0.1259 ± 0.0000	0.1235 ± 0.0000	32.0000 ± 0.0000
	DFS	0.1236	0.1231 ± 0.0006	0.1231 ± 0.0006	**0.0000 ± 0.0000**
	BFT	**0.1264**	**0.1261 ± 0.0003**	**0.1260 ± 0.0003**	1.6154 ± 1.3868
Tmovie	RND	0.3210	0.3176 ± 0.0024	0.3175 ± 0.0024	**0.4118 ± 0.7952**
	FAS	0.3193	0.3192 ± 0.0001	0.3147 ± 0.0000	31.0000 ± 0.0000
	DFS	0.3192	0.3192 ± 0.0000	0.3150 ± 0.0000	33.0000 ± 0.0000
	BFT	**0.3245**	**0.3202 ± 0.0024**	**0.3196 ± 0.0022**	2.5455 ± 1.9164
Cwebkb	RND	0.1341	0.1336 ± 0.0005	0.1335 ± 0.0003	**0.6667 ± 1.1547**
	FAS	0.1321	0.1321 ± 0.0000	0.1307 ± 0.0000	9.0000 ± 0.0000
	DFS	0.1320	0.1320 ± 0.0000	0.1314 ± 0.0002	3.0000 ± 1.4142
	BFT	**0.1341**	**0.1341 ± 0.0000**	**0.1341 ± 0.0000**	1.0000 ± 1.4142
Cr52	RND	0.1775	0.1761 ± 0.0015	0.1761 ± 0.0015	**0.5000 ± 0.5774**
	FAS	0.1750	0.1750 ± 0.0000	0.1740 ± 0.0000	7.0000 ± 0.0000
	DFS	0.1772	**0.1772 ± 0.0000**	0.1763 ± 0.0000	6.0000 ± 0.0000
	BFT	**0.1792**	0.1772 ± 0.0014	**0.1771 ± 0.0014**	0.7500 ± 0.5000
C20ng	RND	0.0917	0.0912 ± 0.0006	0.0912 ± 0.0006	**0.2500 ± 0.5000**
	FAS	0.0905	0.0905 ± 0.0000	0.0897 ± 0.0000	8.0000 ± 0.0000
	DFS	0.0916	0.0916 ± 0.0000	0.0910 ± 0.0000	7.0000 ± 0.0000
	BFT	**0.0930**	**0.0923 ± 0.0004**	**0.0923 ± 0.0004**	0.6000 ± 0.5477
Bbc	RND	0.0835	0.0835 ± 0.0000	0.0835 ± 0.0000	0.0000 ± 0.0000
	FAS	0.0827	0.0826 ± 0.0002	0.0821 ± 0.0001	4.5000 ± 0.7071
	DFS	0.0832	0.0832 ± 0.0000	0.0825 ± 0.0000	5.0000 ± 0.0000
	BFT	**0.0835**	**0.0835 ± 0.0000**	**0.0835 ± 0.0000**	**0.0000 ± 0.0000**
Ad	RND	0.7547	0.7547 ± 0.0000	0.7547 ± 0.0000	**0.0000 ± 0.0000**
	FAS	0.7349	0.7349 ± 0.0000	0.7336 ± 0.0000	1.0000 ± 0.0000
	DFS	0.7290	0.7290 ± 0.0000	0.7235 ± 0.0000	4.0000 ± 0.0000
	BFT	**0.7638**	**0.7601 ± 0.0052**	**0.7600 ± 0.0054**	0.5000 ± 0.7071

Table F.3: *Best score obtained, Average best score obtained, Average initial score generated, Average number of iterations (Avg. It.) using independence selection (best values in bold)*

Appendix G

Empirical Results: Swap Search

Dataset	Approach	Best Score	Avg. Best Score	Avg. Initial Score	Avg. It.
Nltcs	RND	**0.3469**	**0.3440 ± 0.0009**	0.3434 ± 0.0009	500.0000 ± 0.0000
	FAS	0.3469	0.3439 ± 0.0008	0.3436 ± 0.0000	500.0000 ± 0.0000
	DFS	0.3469	0.3440 ± 0.0008	**0.3443 ± 0.0002**	500.0000 ± 0.0000
	BFT	0.3469	0.3440 ± 0.0008	0.3432 ± 0.0008	**500.0000 ± 0.0000**
Msnbc	RND	**0.0828**	**0.0803 ± 0.0011**	0.0789 ± 0.0009	500.0000 ± 0.0000
	FAS	0.0828	0.0802 ± 0.0011	**0.0817 ± 0.0000**	500.0000 ± 0.0000
	DFS	0.0828	0.0800 ± 0.0011	0.0788 ± 0.0003	500.0000 ± 0.0000
	BFT	0.0828	0.0802 ± 0.0010	0.0801 ± 0.0012	**500.0000 ± 0.0000**
Kdd	RND	0.1465	**0.1421 ± 0.0020**	0.1402 ± 0.0008	500.0000 ± 0.0000
	FAS	**0.1497**	0.1421 ± 0.0022	**0.1440 ± 0.0001**	500.0000 ± 0.0000
	DFS	0.1497	0.1421 ± 0.0022	0.1427 ± 0.0004	500.0000 ± 0.0000
	BFT	0.1476	0.1416 ± 0.0020	0.1422 ± 0.0008	**500.0000 ± 0.0000**
Plants	RND	**0.5804**	**0.5732 ± 0.0026**	0.5709 ± 0.0013	194.0682 ± 18.4245
	FAS	0.5799	0.5724 ± 0.0021	0.5726 ± 0.0003	198.6818 ± 20.4516
	DFS	0.5787	0.5727 ± 0.0018	0.5708 ± 0.0006	192.2727 ± 22.9290
	BFT	0.5779	0.5721 ± 0.0022	**0.5740 ± 0.0008**	**188.7045 ± 24.7251**
Baudio	RND	0.1629	0.1554 ± 0.0031	0.1535 ± 0.0010	137.7558 ± 19.8096
	FAS	0.1651	**0.1558 ± 0.0037**	**0.1588 ± 0.0002**	**128.1700 ± 16.1208**
	DFS	**0.1654**	0.1556 ± 0.0036	0.1564 ± 0.0004	135.5000 ± 16.0192
	BFT	0.1642	0.1555 ± 0.0036	0.1567 ± 0.0009	136.7097 ± 24.5219
Bnetflix	RND	0.1085	0.1027 ± 0.0024	0.1013 ± 0.0010	278.1951 ± 54.7107
	FAS	**0.1089**	0.1027 ± 0.0027	**0.1057 ± 0.0002**	**252.8485 ± 42.7771**
	DFS	0.1087	0.1028 ± 0.0027	0.1042 ± 0.0005	259.2857 ± 53.2661
	BFT	0.1086	**0.1034 ± 0.0031**	0.1005 ± 0.0007	278.2903 ± 59.2397
Jester	RND	0.1496	0.1414 ± 0.0027	0.1400 ± 0.0009	68.7000 ± 5.4984
	FAS	0.1491	0.1413 ± 0.0025	**0.1445 ± 0.0004**	64.1900 ± 7.9209
	DFS	**0.1499**	0.1413 ± 0.0025	0.1438 ± 0.0006	**60.6000 ± 6.4307**
	BFT	0.1498	**0.1415 ± 0.0027**	0.1402 ± 0.0007	70.2300 ± 5.3198
Accidents	RND	**0.3819**	0.3584 ± 0.0093	0.3529 ± 0.0043	491.2281 ± 66.2266
	FAS	0.3759	0.3591 ± 0.0110	**0.3668 ± 0.0006**	**470.5882 ± 121.2678**
	DFS	0.3783	**0.3622 ± 0.0105**	0.3662 ± 0.0011	482.7586 ± 92.8477
	BFT	0.3699	0.3566 ± 0.0076	0.3568 ± 0.0036	476.1905 ± 109.1089
Tretail	RND	0.0442	**0.0423 ± 0.0030**	0.0393 ± 0.0012	**333.3333 ± 288.6751**
	FAS	0.0439	0.0416 ± 0.0024	**0.0438 ± 0.0000**	400.0000 ± 223.6068
	DFS	0.0439	0.0416 ± 0.0025	0.0406 ± 0.0001	375.0000 ± 250.0000
	BFT	**0.0443**	0.0418 ± 0.0022	0.0432 ± 0.0012	461.5385 ± 138.6750
Pumsb_star	RND	0.7314	0.6860 ± 0.0131	0.6810 ± 0.0023	29.0000 ± 5.0439
	FAS	**0.7333**	0.6911 ± 0.0115	0.6865 ± 0.0010	24.9700 ± 2.9965
	DFS	0.6932	**0.6926 ± 0.0003**	**0.6903 ± 0.0005**	**22.4900 ± 2.0276**
	BFT	0.7278	0.6853 ± 0.0132	0.6730 ± 0.0027	30.6136 ± 5.1904
Dna	RND	0.1975	0.1961 ± 0.0007	0.1957 ± 0.0003	428.5714 ± 188.9822
	FAS	**0.1984**	0.1966 ± 0.0012	0.1960 ± 0.0002	461.5385 ± 138.6750
	DFS	0.1980	0.1961 ± 0.0013	**0.1990 ± 0.0004**	**428.5714 ± 188.9822**
	BFT	0.1984	**0.1968 ± 0.0011**	0.1971 ± 0.0002	458.3333 ± 144.3376
Kosarek	RND	**0.1817**	**0.1682 ± 0.0038**	0.1643 ± 0.0039	**54.4444 ± 12.0299**
	FAS	0.1805	0.1661 ± 0.0053	0.1789 ± 0.0005	56.3824 ± 12.3067
	DFS	0.1808	0.1671 ± 0.0065	0.1766 ± 0.0014	57.8824 ± 12.6150
	BFT	0.1811	0.1670 ± 0.0063	**0.1791 ± 0.0020**	60.1364 ± 15.1415
Msweb	RND	0.1133	0.1115 ± 0.0025	0.1117 ± 0.0003	359.0000 ± 199.4041
	FAS	0.1171	0.1171 ± 0.0000	**0.1205 ± 0.0000**	323.5000 ± 0.7071
	DFS	**0.1224**	**0.1224 ± 0.0000**	0.1100 ± 0.0003	**188.5000 ± 0.7071**
	BFT	0.1219	0.1189 ± 0.0043	0.1147 ± 0.0000	436.5000 ± 14.8492
Book	RND	**0.1241**	0.1145 ± 0.0015	0.1142 ± 0.0008	6.1100 ± 0.4902
	FAS	0.1224	**0.1216 ± 0.0003**	**0.1213 ± 0.0003**	5.5000 ± 1.0801
	DFS	0.1204	0.1188 ± 0.0006	0.1185 ± 0.0006	**5.3824 ± 1.2064**
	BFT	0.1240	0.1201 ± 0.0026	0.1202 ± 0.0005	5.8710 ± 1.4316
Tmovie	RND	0.2875	0.2856 ± 0.0013	0.2855 ± 0.0013	4.9000 ± 0.9595
	FAS	0.3038	**0.3023 ± 0.0007**	**0.3018 ± 0.0007**	**4.4516 ± 0.9605**
	DFS	0.2960	0.2905 ± 0.0026	0.2898 ± 0.0027	5.2857 ± 1.1501
	BFT	**0.3092**	0.2952 ± 0.0041	0.2949 ± 0.0011	4.6600 ± 0.7416
Cwebkb	RND	0.1251	0.1102 ± 0.0048	0.1087 ± 0.0008	3.6667 ± 1.2309
	FAS	0.1233	**0.1223 ± 0.0005**	**0.1220 ± 0.0005**	3.5625 ± 1.0308
	DFS	0.1215	0.1200 ± 0.0009	0.1198 ± 0.0008	**3.5000 ± 1.2693**
	BFT	**0.1253**	0.1184 ± 0.0038	0.1176 ± 0.0008	3.6818 ± 0.9455
Cr52	RND	0.1607	0.1375 ± 0.0093	0.1352 ± 0.0039	3.2222 ± 1.3017
	FAS	0.1596	**0.1574 ± 0.0012**	**0.1569 ± 0.0012**	3.2273 ± 0.8691
	DFS	0.1568	0.1561 ± 0.0004	0.1557 ± 0.0004	3.0000 ± 1.0000
	BFT	**0.1627**	0.1497 ± 0.0086	0.1494 ± 0.0037	**3.0000 ± 0.8165**
C20ng	RND	0.0697	0.0680 ± 0.0009	0.0680 ± 0.0009	0.9767 ± 0.1525
	FAS	**0.0770**	**0.0763 ± 0.0003**	**0.0763 ± 0.0003**	0.9853 ± 0.1213
	DFS	0.0755	0.0741 ± 0.0005	0.0741 ± 0.0005	**0.9677 ± 0.1796**
	BFT	0.0751	0.0734 ± 0.0007	0.0734 ± 0.0007	0.9863 ± 0.1170
Bbc	RND	**0.0783**	0.0677 ± 0.0060	0.0648 ± 0.0010	6.4000 ± 3.5777
	FAS	0.0766	**0.0763 ± 0.0004**	**0.0760 ± 0.0002**	7.2000 ± 4.0866
	DFS	0.0723	0.0715 ± 0.0006	0.0712 ± 0.0007	6.0000 ± 4.0000
	BFT	0.0714	0.0679 ± 0.0039	0.0716 ± 0.0005	**6.0000 ± 4.0000**
Ad	RND	0.7495	0.7495 ± 0.0000	0.6863 ± 0.0000	**0.5000 ± 0.7071**
	FAS	0.6950	0.6933 ± 0.0023	0.6927 ± 0.0027	4.0000 ± 2.8284
	DFS	0.6983	0.6968 ± 0.0015	**0.6958 ± 0.0015**	4.7500 ± 3.2016
	BFT	**0.7556**	**0.7556 ± 0.0000**	0.6909 ± 0.0000	3.5000 ± 0.7071

Table G.1: *Best score obtained, Average best score obtained, Average initial score generated, Average number of iterations (Avg. It.) using sequential selection (best values in bold)*

Dataset	Approach	Best Score	Avg. Best Score	Avg. Initial Score	Avg. It.
Nltcs	RND	0.3605	0.3366 ± 0.0048	0.3325 ± 0.0030	500.0000 ± 0.0000
	FAS	0.3605	0.3368 ± 0.0048	**0.3376 ± 0.0010**	500.0000 ± 0.0000
	DFS	0.3605	0.3363 ± 0.0051	0.3374 ± 0.0007	500.0000 ± 0.0000
	BFT	**0.3643**	**0.3377 ± 0.0053**	0.3318 ± 0.0030	**500.0000 ± 0.0000**
Msnbc	RND	0.0603	0.0528 ± 0.0032	0.0496 ± 0.0033	500.0000 ± 0.0000
	FAS	0.0603	0.0527 ± 0.0034	0.0540 ± 0.0000	500.0000 ± 0.0000
	DFS	0.0603	0.0527 ± 0.0034	**0.0543 ± 0.0005**	500.0000 ± 0.0000
	BFT	**0.0615**	**0.0529 ± 0.0033**	0.0517 ± 0.0030	**500.0000 ± 0.0000**
Kdd	RND	0.1452	0.1379 ± 0.0030	0.1353 ± 0.0011	500.0000 ± 0.0000
	FAS	0.1449	0.1383 ± 0.0033	**0.1389 ± 0.0002**	500.0000 ± 0.0000
	DFS	0.1470	0.1376 ± 0.0032	0.1371 ± 0.0002	500.0000 ± 0.0000
	BFT	**0.1492**	**0.1385 ± 0.0034**	0.1374 ± 0.0014	**500.0000 ± 0.0000**
Plants	RND	0.5838	0.5637 ± 0.0072	0.5566 ± 0.0024	500.0000 ± 0.0000
	FAS	0.5772	0.5632 ± 0.0078	**0.5615 ± 0.0007**	500.0000 ± 0.0000
	DFS	**0.5862**	0.5635 ± 0.0085	0.5575 ± 0.0009	500.0000 ± 0.0000
	BFT	0.5829	**0.5650 ± 0.0091**	0.5588 ± 0.0019	**486.4865 ± 82.1995**
Baudio	RND	0.1651	**0.1578 ± 0.0046**	0.1545 ± 0.0009	456.2667 ± 127.5032
	FAS	**0.1658**	0.1563 ± 0.0041	0.1567 ± 0.0005	**447.6842 ± 112.1378**
	DFS	0.1652	0.1564 ± 0.0043	**0.1575 ± 0.0003**	459.3704 ± 95.8385
	BFT	0.1648	0.1572 ± 0.0042	0.1565 ± 0.0010	447.8333 ± 142.9621
Bnetflix	RND	0.1069	0.1018 ± 0.0024	0.1009 ± 0.0007	450.0000 ± 158.1139
	FAS	**0.1094**	0.1034 ± 0.0038	0.1022 ± 0.0003	**450.0000 ± 158.1139**
	DFS	0.1088	0.1018 ± 0.0032	**0.1037 ± 0.0005**	454.5455 ± 150.7557
	BFT	0.1094	**0.1035 ± 0.0037**	0.1002 ± 0.0007	464.2857 ± 133.6306
Jester	RND	0.1411	0.1394 ± 0.0010	0.1389 ± 0.0011	361.2308 ± 111.3494
	FAS	0.1472	**0.1409 ± 0.0025**	**0.1446 ± 0.0003**	357.7222 ± 98.7152
	DFS	0.1418	0.1397 ± 0.0013	0.1431 ± 0.0006	362.3043 ± 87.2454
	BFT	**0.1477**	0.1405 ± 0.0029	0.1397 ± 0.0008	**345.6667 ± 101.4260**
Accidents	RND	0.3847	0.3686 ± 0.0071	0.3644 ± 0.0036	479.1667 ± 102.0621
	FAS	0.3914	0.3689 ± 0.0100	0.3704 ± 0.0011	472.2222 ± 117.8511
	DFS	**0.3925**	0.3666 ± 0.0104	**0.3735 ± 0.0004**	**458.3333 ± 144.3376**
	BFT	0.3876	**0.3720 ± 0.0092**	0.3642 ± 0.0039	466.6667 ± 129.0994
Tretail	RND	**0.0442**	0.0425 ± 0.0016	0.0411 ± 0.0004	333.3333 ± 288.6751
	FAS	0.0440	0.0430 ± 0.0007	0.0430 ± 0.0000	444.4444 ± 166.6667
	DFS	0.0437	**0.0436 ± 0.0002**	0.0423 ± 0.0002	250.0000 ± 353.5534
	BFT	0.0427	0.0424 ± 0.0004	**0.0442 ± 0.0000**	**250.0000 ± 353.5534**
Pumsb_star	RND	0.7047	**0.6676 ± 0.0254**	0.6542 ± 0.0047	416.6667 ± 204.1241
	FAS	0.6549	0.6506 ± 0.0043	0.6675 ± 0.0011	**400.0000 ± 223.6068**
	DFS	0.7087	0.6656 ± 0.0245	**0.6704 ± 0.0023**	444.4444 ± 166.6667
	BFT	**0.7126**	0.6656 ± 0.0244	0.6495 ± 0.0050	470.5882 ± 121.2678
Dna	RND	0.1981	0.1961 ± 0.0019	0.1944 ± 0.0008	416.6667 ± 204.1241
	FAS	**0.1984**	**0.1967 ± 0.0021**	0.1953 ± 0.0001	333.3333 ± 288.6751
	DFS	0.1982	0.1959 ± 0.0022	**0.1990 ± 0.0003**	333.3333 ± 288.6751
	BFT	0.1978	0.1955 ± 0.0021	0.1957 ± 0.0010	**333.3333 ± 288.6751**
Kosarek	RND	0.1775	0.1676 ± 0.0068	0.1683 ± 0.0033	**295.7500 ± 202.7303**
	FAS	0.1792	0.1701 ± 0.0050	0.1706 ± 0.0004	345.6000 ± 126.3515
	DFS	0.1681	0.1673 ± 0.0010	0.1665 ± 0.0005	338.2000 ± 191.1889
	BFT	**0.1797**	**0.1730 ± 0.0065**	**0.1784 ± 0.0014**	320.0000 ± 181.3464
Msweb	RND	0.1216	**0.1216 ± 0.0000**	0.1104 ± 0.0000	254.5000 ± 0.7071
	FAS	0.1187	0.1149 ± 0.0053	**0.1170 ± 0.0001**	317.0000 ± 258.8011
	DFS	0.1127	0.1127 ± 0.0000	0.1089 ± 0.0000	**242.5000 ± 0.7071**
	BFT	**0.1219**	0.1205 ± 0.0019	0.1159 ± 0.0027	317.5000 ± 258.0940
Book	RND	0.1279	0.1204 ± 0.0042	0.1186 ± 0.0005	27.0000 ± 15.1822
	FAS	0.1272	**0.1265 ± 0.0003**	**0.1258 ± 0.0003**	**26.8421 ± 7.6031**
	DFS	0.1195	0.1190 ± 0.0003	0.1183 ± 0.0003	28.3333 ± 9.0988
	BFT	**0.1283**	0.1203 ± 0.0036	0.1245 ± 0.0007	27.4286 ± 12.5546
Tmovie	RND	**0.3389**	0.3156 ± 0.0094	0.3114 ± 0.0017	17.4444 ± 3.5554
	FAS	0.3274	**0.3258 ± 0.0009**	**0.3245 ± 0.0009**	13.1719 ± 2.1199
	DFS	0.3111	0.3074 ± 0.0018	0.3062 ± 0.0018	16.1500 ± 4.2212
	BFT	0.3347	0.3203 ± 0.0084	0.3231 ± 0.0012	**12.9655 ± 3.1223**
Cwebkb	RND	0.1245	0.1238 ± 0.0008	0.1239 ± 0.0005	6.6667 ± 5.7735
	FAS	0.1336	0.1333 ± 0.0004	**0.1330 ± 0.0003**	7.2500 ± 4.8563
	DFS	0.1217	0.1211 ± 0.0005	0.1207 ± 0.0005	6.7500 ± 4.7170
	BFT	**0.1385**	**0.1350 ± 0.0030**	0.1326 ± 0.0009	**6.6000 ± 3.7148**
Cr52	RND	**0.1851**	0.1676 ± 0.0120	0.1616 ± 0.0030	5.7500 ± 3.8622
	FAS	0.1790	**0.1787 ± 0.0003**	**0.1781 ± 0.0002**	5.8000 ± 3.3466
	DFS	0.1498	0.1480 ± 0.0019	0.1473 ± 0.0019	**5.3333 ± 4.6188**
	BFT	0.1766	0.1720 ± 0.0055	0.1731 ± 0.0022	6.1000 ± 2.2336
C20ng	RND	0.0872	0.0855 ± 0.0010	0.0854 ± 0.0010	2.7619 ± 0.7003
	FAS	0.0943	**0.0939 ± 0.0002**	**0.0938 ± 0.0002**	**2.7500 ± 0.7746**
	DFS	0.0849	0.0844 ± 0.0004	0.0843 ± 0.0004	2.7500 ± 0.8660
	BFT	**0.0946**	0.0937 ± 0.0005	0.0937 ± 0.0005	2.8000 ± 0.7746
Bbc	RND	0.0841	0.0791 ± 0.0056	0.0736 ± 0.0007	6.2500 ± 4.1932
	FAS	0.0803	**0.0801 ± 0.0002**	**0.0798 ± 0.0001**	**6.0000 ± 4.0000**
	DFS	0.0739	0.0732 ± 0.0006	0.0730 ± 0.0005	6.0000 ± 5.1962
	BFT	**0.0842**	0.0774 ± 0.0060	0.0757 ± 0.0005	6.3333 ± 5.5076
Ad	RND	**0.7694**	**0.7694 ± 0.0000**	0.7055 ± 0.0000	**1.5000 ± 0.7071**
	FAS	0.7121	0.7111 ± 0.0014	0.7104 ± 0.0012	6.0000 ± 0.0000
	DFS	0.7148	0.7148 ± 0.0000	**0.7141 ± 0.0004**	3.5000 ± 0.7071
	BFT	0.7083	0.7078 ± 0.0007	0.7081 ± 0.0001	5.0000 ± 0.0000

Table G.2: *Best score obtained, Average best score obtained, Average initial score generated, Average number of iterations (Avg. It.) using greedy selection (best values in bold)*

Dataset	Approach	Best Score	Avg. Best Score	Avg. Initial Score	Avg. It.
Nltcs	RND	**0.3480**	**0.3440 ± 0.0008**	0.3433 ± 0.0010	500.0000 ± 0.0000
	FAS	0.3480	0.3440 ± 0.0009	0.3436 ± 0.0000	500.0000 ± 0.0000
	DFS	0.3480	0.3440 ± 0.0008	**0.3443 ± 0.0002**	500.0000 ± 0.0000
	BFT	0.3480	0.3440 ± 0.0009	0.3432 ± 0.0008	**500.0000 ± 0.0000**
Msnbc	RND	**0.0824**	0.0783 ± 0.0022	0.0751 ± 0.0037	500.0000 ± 0.0000
	FAS	0.0824	0.0780 ± 0.0028	**0.0771 ± 0.0000**	500.0000 ± 0.0000
	DFS	0.0824	**0.0785 ± 0.0021**	0.0759 ± 0.0014	500.0000 ± 0.0000
	BFT	0.0824	0.0782 ± 0.0023	0.0738 ± 0.0039	**500.0000 ± 0.0000**
Kdd	RND	0.1387	0.1286 ± 0.0049	0.1247 ± 0.0029	500.0000 ± 0.0000
	FAS	0.1389	0.1295 ± 0.0055	0.1230 ± 0.0005	500.0000 ± 0.0000
	DFS	0.1381	0.1287 ± 0.0046	0.1181 ± 0.0002	500.0000 ± 0.0000
	BFT	**0.1414**	**0.1299 ± 0.0055**	**0.1268 ± 0.0028**	**500.0000 ± 0.0000**
Plants	RND	0.5610	0.5330 ± 0.0141	0.5235 ± 0.0041	476.1905 ± 109.1089
	FAS	0.5653	0.5348 ± 0.0156	0.5214 ± 0.0021	473.6842 ± 114.7079
	DFS	**0.5674**	**0.5359 ± 0.0156**	**0.5355 ± 0.0018**	**470.5882 ± 121.2678**
	BFT	0.5618	0.5352 ± 0.0159	0.5246 ± 0.0051	476.1905 ± 109.1089
Baudio	RND	0.1470	**0.1364 ± 0.0073**	0.1292 ± 0.0030	482.1429 ± 94.4911
	FAS	0.1465	0.1347 ± 0.0088	0.1310 ± 0.0003	468.7500 ± 125.0000
	DFS	0.1463	0.1364 ± 0.0074	0.1215 ± 0.0013	**468.7500 ± 125.0000**
	BFT	**0.1492**	0.1336 ± 0.0065	**0.1375 ± 0.0023**	489.3617 ± 72.9325
Bnetflix	RND	0.0988	0.0897 ± 0.0050	0.0871 ± 0.0026	477.2727 ± 106.6004
	FAS	**0.0998**	0.0907 ± 0.0056	0.0872 ± 0.0002	483.8710 ± 89.8027
	DFS	0.0971	**0.0912 ± 0.0042**	0.0806 ± 0.0013	472.2222 ± 117.8511
	BFT	0.0977	0.0896 ± 0.0045	**0.0891 ± 0.0021**	**464.2857 ± 133.6306**
Jester	RND	0.1341	**0.1216 ± 0.0088**	0.1149 ± 0.0036	**487.1795 ± 80.0641**
	FAS	**0.1342**	0.1184 ± 0.0074	**0.1201 ± 0.0004**	493.0556 ± 58.9256
	DFS	0.1333	0.1207 ± 0.0079	0.1080 ± 0.0012	493.2432 ± 58.1238
	BFT	0.1333	0.1179 ± 0.0068	0.1197 ± 0.0026	493.0556 ± 58.9256
Accidents	RND	0.3505	**0.3147 ± 0.0190**	0.2978 ± 0.0083	472.2222 ± 117.8511
	FAS	0.3508	0.3099 ± 0.0191	**0.3204 ± 0.0006**	**464.2857 ± 133.6306**
	DFS	0.3378	0.3039 ± 0.0132	0.3044 ± 0.0013	480.7692 ± 98.0581
	BFT	**0.3582**	0.3079 ± 0.0185	0.3046 ± 0.0068	491.6667 ± 64.5497
Tretail	RND	0.0441	0.0437 ± 0.0007	0.0386 ± 0.0020	333.3333 ± 288.6751
	FAS	**0.0442**	0.0435 ± 0.0009	**0.0436 ± 0.0000**	375.0000 ± 250.0000
	DFS	0.0440	**0.0440 ± 0.0000**	0.0366 ± 0.0001	**250.0000 ± 353.5534**
	BFT	0.0441	0.0400 ± 0.0029	0.0435 ± 0.0007	416.6667 ± 204.1241
Pumsb_star	RND	0.6633	0.6224 ± 0.0274	0.6064 ± 0.0099	375.0000 ± 250.0000
	FAS	0.6682	**0.6396 ± 0.0352**	0.6299 ± 0.0017	400.0000 ± 223.6068
	DFS	0.6785	0.6271 ± 0.0449	**0.6349 ± 0.0020**	**333.3333 ± 288.6751**
	BFT	**0.6795**	0.6312 ± 0.0382	0.5981 ± 0.0042	428.5714 ± 188.9822
Dna	RND	0.1927	0.1909 ± 0.0027	0.1885 ± 0.0003	250.0000 ± 353.5534
	FAS	0.1892	0.1889 ± 0.0004	**0.1945 ± 0.0000**	250.0000 ± 353.5534
	DFS	**0.1932**	**0.1924 ± 0.0011**	0.1919 ± 0.0005	250.0000 ± 353.5534
	BFT	0.1920	0.1904 ± 0.0022	0.1902 ± 0.0004	**250.0000 ± 353.5534**
Kosarek	RND	0.1582	0.1555 ± 0.0019	0.1576 ± 0.0051	**375.0000 ± 250.0000**
	FAS	0.1769	0.1636 ± 0.0097	0.1698 ± 0.0010	437.5000 ± 176.7767
	DFS	0.1763	**0.1653 ± 0.0087**	0.1520 ± 0.0008	416.6667 ± 204.1241
	BFT	**0.1778**	0.1614 ± 0.0096	**0.1717 ± 0.0032**	464.2857 ± 133.6306
Msweb	RND	0.1103	0.1093 ± 0.0013	0.1033 ± 0.0018	**250.0000 ± 353.5534**
	FAS	0.1079	0.1054 ± 0.0043	**0.1166 ± 0.0000**	333.3333 ± 288.6751
	DFS	0.1164	0.1164 ± 0.0000	0.0930 ± 0.0018	438.5000 ± 0.7071
	BFT	**0.1189**	**0.1189 ± 0.0000**	0.1131 ± 0.0003	304.5000 ± 0.7071
Book	RND	0.1256	0.1178 ± 0.0068	0.1148 ± 0.0012	63.3333 ± 54.8574
	FAS	0.1253	**0.1218 ± 0.0051**	0.1211 ± 0.0002	92.0000 ± 61.3406
	DFS	0.1253	0.1163 ± 0.0055	0.1111 ± 0.0009	92.5000 ± 37.3898
	BFT	**0.1257**	0.1194 ± 0.0055	**0.1218 ± 0.0005**	**49.0000 ± 42.4617**
Tmovie	RND	**0.3170**	**0.3015 ± 0.0219**	0.2732 ± 0.0066	**41.5000 ± 58.6899**
	FAS	0.3159	0.2917 ± 0.0177	0.2875 ± 0.0006	76.0000 ± 50.6952
	DFS	0.2803	0.2699 ± 0.0095	0.2486 ± 0.0043	84.8333 ± 42.8692
	BFT	0.2787	0.2763 ± 0.0027	**0.2998 ± 0.0066**	55.0000 ± 47.6340
Cwebkb	RND	0.1180	0.1160 ± 0.0029	0.1158 ± 0.0027	**12.0000 ± 15.5563**
	FAS	**0.1271**	**0.1266 ± 0.0006**	0.1246 ± 0.0002	14.6667 ± 12.7410
	DFS	0.1130	0.1125 ± 0.0007	0.1110 ± 0.0005	19.5000 ± 10.6066
	BFT	0.1177	0.1177 ± 0.0000	**0.1269 ± 0.0000**	21.5000 ± 0.7071
Cr52	RND	**0.1753**	0.1579 ± 0.0246	0.1475 ± 0.0005	9.5000 ± 13.4350
	FAS	0.1660	**0.1653 ± 0.0006**	0.1629 ± 0.0003	15.7500 ± 10.6262
	DFS	0.1336	0.1326 ± 0.0009	0.1298 ± 0.0010	14.0000 ± 12.1655
	BFT	0.1753	0.1557 ± 0.0276	**0.1636 ± 0.0008**	**9.0000 ± 12.7279**
C20ng	RND	0.0724	0.0722 ± 0.0003	0.0705 ± 0.0020	6.0000 ± 8.4853
	FAS	0.0834	**0.0832 ± 0.0002**	0.0822 ± 0.0001	11.4000 ± 6.4653
	DFS	0.0602	0.0600 ± 0.0003	0.0589 ± 0.0002	10.6667 ± 9.2376
	BFT	**0.0870**	0.0799 ± 0.0099	**0.0854 ± 0.0019**	**6.0000 ± 8.4853**
Bbc	RND	0.0826	0.0793 ± 0.0058	0.0705 ± 0.0018	8.0000 ± 6.9282
	FAS	0.0794	0.0794 ± 0.0000	**0.0790 ± 0.0001**	**6.5000 ± 0.7071**
	DFS	0.0725	0.0717 ± 0.0008	0.0713 ± 0.0007	9.3333 ± 8.0829
	BFT	**0.0836**	**0.0836 ± 0.0000**	0.0740 ± 0.0000	10.5000 ± 0.7071
Ad	RND	**0.7563**	0.7263 ± 0.0424	0.6926 ± 0.0035	4.5000 ± 0.7071
	FAS	0.7057	0.7034 ± 0.0022	**0.7017 ± 0.0023**	4.6667 ± 4.0415
	DFS	0.6993	0.6987 ± 0.0008	0.6973 ± 0.0020	**4.0000 ± 4.2426**
	BFT	0.7557	**0.7557 ± 0.0000**	0.6958 ± 0.0000	4.5000 ± 0.7071

Table G.3: *Best score obtained, Average best score obtained, Average initial score generated, Average number of iterations (Avg. It.) using independence selection (best values in bold)*

Bibliography

[ACMG13] Alessandro Antonucci, Giorgio Corani, Denis D. Mauá e Sandra Gabaglio. An ensemble of Bayesian networks for multilabel classification. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, páginas 1220–1225, 2013. 33

[Aka74] Hirotugu Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974. 1

[BC15] Mark Bartlett e James Cussens. Integer linear programming for the Bayesian network structure learning problem. *Artificial Intelligence*, 2015. 7, 8, 10

[Bun91] Wray Buntine. Theory refinement on Bayesian networks. In *Proceedings of the 7th Annual Conference on Uncertainty Artificial Intelligence*, páginas 52–60, 1991. 2

[CG99] Jie Cheng e Russell Greiner. Comparing Bayesian network classifiers. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, páginas 101–108, 1999. 33

[CGK+02] Jie Cheng, Russell Greiner, Jonathan Kelly, David Bell e Weiru Liu. Learning Bayesian networks from data: An information-theory based approach. *Artificial Intelligence*, 137(1-2):43–90, 2002. 1

[CH92] Gregory F. Cooper e Edward Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9(4):309–347, 1992. 1, 9

[Chi96] David M. Chickering. Learning Bayesian networks is NP-Complete. In *Learning from Data: Artificial Intelligence and Statistics V*, páginas 121–130, 1996. 1

[Chi02] David M. Chickering. Learning equivalence classes of Bayesian-network structures. *Journal of Machine Learning Research*, 2:445–498, 2002. 1

[CHM04] David M. Chickering, David Heckerman e Christopher Meek. Large-sample learning of Bayesian networks is NP-Hard. *Journal of Machine Learning Research*, 5(1):1287–1330, 2004. 1, 5

[CT91] Thomas M. Cover e Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 1991. 1

[Dar09] Adnan Darwiche. *Modeling and reasoning with Bayesian Networks*. Cambridge University Press, 2009. 1

[dCJ11] Cassio P. de Campos e Qiang Ji. Efficient structure learning of Bayesian networks using constraints. *Journal of Machine Learning Research*, 12:663–689, 2011. 6, 7, 8

[dCZJ09] Cassio P. de Campos, Zhi Zeng e Qiang Ji. Structure learning of Bayesian networks using constraints. In *Proceedings of the 26st Annual International Conference on Machine Learning*, páginas 113–120, 2009. 7

[Dem06] Janez Demsar. Statistical comparison of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006. 25

[DF03] Camil Demetrescu e Irene Finocchi. Combinatorial algorithms for feedback problems in directed graphs. *Information Processing Letters*, 86(3):129–136, 2003. 19

[dWG07] Peter de Waal e Linda C. Van Der Gaag. Inference and learning in multi-dimensional Bayesian network classifiers. *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, páginas 501–511, 2007. 33

[EL95] P. Eades e X. Lin. A new heuristic for the feedback arc set problem. *Australian Journal of Combinatorics*, 12:15–26, 1995. 19

[ENFS02] Gal Elidan, Matan Ninio, Nir Friedman e Dale Schuurmans. Data perturbation for escaping local maxima in learning. In *Eighteenth National Conference on Artificial Intelligence*, páginas 132–139, 2002. 2

[EXS93] P. Eades, X.Lin e W. F. Smyth. A fast and effective heuristic for the feedback arc set problem. *Information Processing Letters*, 47(6):319–323, 1993. 19

[FG96] Nir Friedman e Moises Goldszmidt. Building classifiers using Bayesian networks. In *Proceedings of the National Conference on Artificial Intelligence*, páginas 1277–1284, 1996. 33

[FNP99] Nir Friedman, Iftach Nachman e Dana Peér. Learning Bayesian network structure from massive datasets: The "sparse candidate" algorithm. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, páginas 206–215, 1999. 1

[Gav77] F. Gavril. Some NP-complete problems on graphs. In *Proceedings of the 11th Conference on Information Sciences and Systems*, páginas 91–95, 1977. 19

[GKR94] Vincent Granville, Mirko Krivanek e Jean Paul Rasson. Simulated annealing: A proof of convergence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6):652–656, 1994. 2, 11

[Glo89] Fred Glover. Tabu search - part I. *Operations Research Society of America Journal on Computing*, 1(3):190–206, 1989. 2, 12

[HGC95] David Heckerman, Dan Geiger e David M. Chickering. Learning Bayesian networks: The combination of knowledge and statistical data. *Journal of Machine Learning Research*, 20(MSR-TR-94-09):197–243, 1995. 1

[HMC06] David Heckerman, Cristopher Meek e Gregory Cooper. *Innovations in Machine Learning: Theory and Applications*, chapter A Bayesian Approach to Causal Discovery, páginas 1–28. Springer Berlin Heidelberg, 2006. 1

[Knu98] Donald Knuth. *The Art of Computer Programming 2*. Boston: Adison-Wesley, 1998. 17

[Koi06] Mikko Koivisto. Parent assignment is hard for the MDL, AIC, and NML costs. In *Proceedings of the 19st Annual Conference on Learning Theory*, páginas 289–303, 2006. 6

[LB94] Wai Lam e Fahiem Bacchus. Learning Bayesian belief networks: An approach based on the MDL principle. *Computational Intelligence*, 10(4):31, 1994. 1

[Mad08] Michael G. Madden. On the classification performance of TAN and general Bayesian networks. *Knowledge-Based Systems*, 22(7):489–495, 2008. 33

[Mal12] Brandon M. Malone. *Learning optimal Bayesian networks with heuristic search.* Tese de Doutorado, Mississippi State University, Mississipi, USA, 2012. 21

[McG54] William J. McGill. Multivariate information transmission. *Psychometrika*, 19(2):97–116, 1954. 9

[ML98] Andrew Moore e Mary Soon Lee. Cached sufficient statistics for efficient machine learning with large datasets. *Journal of Artificial Intelligence*, 8:67–91, 1998. 10

[Pea88] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, 1988. 1

[PIM08] Eric Perrier, Seiya Imoto e Satoru Miyano. Finding optimal Bayesian network given a super-structure. *J. of Mach. Learning Res.*, 9(2):2251–2286, 2008. 10

[PM15] Walter Perez e Denis D. Mauá. Initialization heuristics for greedy Bayesian network structure learning. In *Proceedings of the 3rd Symposium on Knowledge Discovery, Mining and Learning*, páginas 58–65, 2015. 2

[PM16] Walter Perez e Denis Mauá. Improving acyclic selection order-based Bayesian network structure learning. In *Proceedings of the XIII Encontro Nacional de Inteligência Artificial*, 2016. 2, 20, 29

[Rob73] Robert W. Robinson. Counting labeled acyclic digraphs. *Harary, F., New Directions in the Theory of Graphs*, páginas 239–273, 1973. 6

[Sch78] Gideon Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978. 1

[SdCCZ15] Mauro Scanagatta, Cassio P. de Campos, Giorgio Corani e Marco Zaffalon. Learning Bayesian networks with thousands of variables. In *Proceedings of the 29th Annual Conference on Neural Information Processing Systems*, páginas 1855–1863, 2015. 2, 6, 7, 9, 13, 14, 23

[SM95] Peter Spirtes e Cristopher Meek. Learning Bayesian networks with discrete variables from data. In *Proceedings of the 1st International Conference on Knowledge Discovery and Data Mining*, página 294–299, 1995. 1

[TBA06] Ioannis Tsamardinos, Laura E. Brown e Constantin F. Aliferis. The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning*, 65(1):31–78, 2006. 1

[TK05] Marc Teyssier e Daphne Koller. Ordering-based search: A simple and effective algorithm for learning Bayesian networks. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence, UAI 2005*, páginas 584–590, 2005. 2

[YM12] Changhe Yuan e Brandon Malone. An improved admissible heuristic for learning optimal Bayesian networks. In *Proceddings of the 28th Conference on Uncertainty in Artificial Intelligence*, páginas 924–933, 2012. 10

[YM13] Changhe Yuan e Brandon Malone. Learning optimal Bayesian networks: A shortest path perspective. *Artificial Intelligence*, 48:23–65, 2013. 7, 8

[YMW11] Changhe Yuan, Brandon Malone e Xiaojian Wu. Learning optimal Bayesian networks using A* search. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, páginas 2186–2191, 2011. 10