



HAL
open science

On the Global Structure of PUBOi Fitness Landscapes

Sara Tari, Gabriela Ochoa, Matthieu Basseur, Sébastien Verel

► **To cite this version:**

Sara Tari, Gabriela Ochoa, Matthieu Basseur, Sébastien Verel. On the Global Structure of PUBOi Fitness Landscapes. The Genetic and Evolutionary Computation Conference (GECCO), Jul 2023, Lisbonne, Portugal. pp.247-250, 10.1145/3583133.3590649 . hal-04075074

HAL Id: hal-04075074

<https://hal.science/hal-04075074v1>

Submitted on 19 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the Global Structure of PUBO_i Fitness Landscapes

Sara Tari

Univ. Littoral Côte d'Opale
LISIC, F-62100 Calais, France
sara.tari@univ-littoral.fr

Matthieu Basseur

Univ. Littoral Côte d'Opale
LISIC, F-62100 Calais, France
matthieu.basseur@univ-littoral.fr

Gabriela Ochoa

University of Stirling
Stirling, Scotland, United Kingdom
gabriela.ochoa@stir.ac.uk

Sebastien Verel

Univ. Littoral Côte d'Opale
LISIC, F-62100 Calais, France
verel@univ-littoral.fr

ABSTRACT

In most of the existing benchmark generators for combinatorial optimization, one cannot tune variable importance, making the generation of real-like instances challenging. However, when it comes to the study of optimization algorithms or new problems there is a lack of real-like instances. To achieve more real-like instances, a recently proposed generator, PUBO_i (Polynomial Unconstrained Binary Optimization with importance), includes parameters that directly influence variable weights. The parameters of this generator enable the generation of landscapes with varying properties, such as ruggedness and neutrality levels, yet their global structure and the impact on optimization algorithm behavior remain to be studied. In this work, we use local optima networks to observe the differences in the landscape's global structure according to variable importance. Both the visualization and the metrics highlight how the landscapes are affected by the variable importance parameters of PUBO_i, and that landscapes with variable importance resemble real-like ones previously observed. We also conduct a first performance analysis using two iterated local search algorithms and observe different behaviors on different PUBO_i instances according to variable weights distribution.

CCS CONCEPTS

• **Mathematics of computing** → **Graph algorithms**; *Combinatorial algorithms*; • **Theory of computation** → **Evolutionary algorithms**.

KEYWORDS

benchmarking, combinatorial optimization, local optima networks

ACM Reference Format:

Sara Tari, Gabriela Ochoa, Matthieu Basseur, and Sebastien Verel. 2023. On the Global Structure of PUBO_i Fitness Landscapes. In *Genetic and Evolutionary Computation Conference Companion (GECCO '23 Companion)*, July 15–19, 2023, Lisbon, Portugal. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3583133.3590649>

1 INTRODUCTION

The PUBO_i (Polynomial Unconstrained Binary Optimization with importance) generator [14] was introduced recently and allows the generation of pseudo-Boolean optimization problems with configurable variable importance. A first analysis highlighted the impact of benchmark parameters on some structural properties of the fitness landscapes stemming from these instances, namely their ruggedness and neutrality. While such characteristics are relevant and widely considered, they do not provide much insight into the fitness landscape global structure, which is known to highly impact the optimization quality of evolutionary algorithms.

In this work, we propose a study of various PUBO_i fitness landscapes by (1) considering local optima networks (LONs) to visualize their global structures and investigate the impact of the benchmark parameters on LONs metrics, (2) providing the first performance analysis of optimization algorithms on PUBO_i benchmarks. The latter uses an iterated local search method with two perturbation operators and mostly focuses on a parameter of PUBO_i called the co-appearance factor.

The paper is organized as follows. Section 2 provides relevant definitions and overviews previous related work. Section 3 presents the global structure analysis and visualization of PUBO_i fitness landscapes. A performance analysis of iterated local search algorithms is provided in section 3, with a particular focus on the variable importance parameters across PUBO_i instances. Section 5 summarizes our main findings and points out possible directions forward.

2 BACKGROUND

2.1 PUBO Problems

Quadratic Unconstrained Binary Optimization (QUBO) are well-known pseudo-Boolean functions in the field of physics [2] that can be generalized to any order. They are known under several names, for example, Unconstrained Binary Quadratic Problem (UBQP) [6], or PUBO for Polynomial Unconstrained Binary optimization [4]. In the following, we refer to such problems as PUBO and discuss the links with Walsh functions [16].

Walsh functions enable the representation of any evaluation function as well as the fine-tuning of variables interdependency (non-zero terms of the polynomial) and the interaction intensity ($|w_k|$ values). While different generators exist for specific PUBO problems, none of them is expressed using directly Walsh functions. The closest to this idea is a python-based generator for the Tile Planting (TP) problem called *Chook*, proposed by Perera *et al.* [12] in the field of physics. TP problems correspond to PUBO problems, in

which global optima are known (planting solutions). In most cases such problems can be expressed using only 2-local interactions between variables (spins) with a graph.

In TP instances, the graph $G = (V, E)$ is decomposed into edge-disjoint and vertex-sharing subgraphs $\{G_l = (V_l, E_l)\}$ [5]. To each subgraph corresponds a function allowing the TP problems to be expressed as their sum. TP problems use a regular lattice structure with different variants that enables the decomposition of the problem graph into subgraphs containing a subset of the unit cells. In the square lattice variant, the problem graph corresponds to a toric square matrix, in which unit cells form a checkboard pattern. Chook defines a portfolio of four sub-function classes, in which a function class C_j has j local minima, individually, the higher the j the harder to solve the sub-problem. The instance generation process in Chook assigns a sub-function to each problem subgraph following a probability distribution, and randomly rotates the plaqette in the lattice. The probability distribution related to each sub-problem influences the problem difficulty, yet the 2 dimension shape of variable interactions makes solving these instances possible in a polynomial time [3]. Another limitation is that square toric lattices give similar weight to each binary variable, leading to benchmarks that differ from most real-world problems.

2.2 PUBO_i Generator

The generator PUBO_i (Polynomial Unconstrained Binary Optimization with importance) was introduced in [14] with the aim to generate more structured benchmarks as in many real-life problems. PUBO_i benchmarks use the Chook portfolio of sub-problems and three parameters that influence variable importance.

As in TP instances, for n variables the objective is defined as a sum of sub-functions: $\forall x \in \{0, 1\}^n$, $f(x) = \sum_{i=1}^m f_i(x)$, with m the number of sub-functions. Each sub-function f_i is randomly selected according to probabilities p_j of each sub-function class C_j from the portfolio. Each sub-function depends on four variables selected following a degree of importance.

In PUBO_i variables are split into k disjunctive classes of importance c_i , each composed of n_i variables and having a degree of importance $d_i \in \mathbb{R}^+$. For each class i , the degree of importance d_i determines the probability of its variables to be selected in a sub-function as follows: $p_{c_i} = \frac{d_i}{\sum_{j=1}^k d_j}$.

Another parameter of the generator is α the co-appearance factor of important variables that affect the distribution of important variables among the benchmark sub-problems. $\alpha = 1$ leads to an independent co-appearance of variables importance class, while $\alpha > 1$ leads to a higher probability of having the same class variables in the same sub-function, and $\alpha < 1$ a lower probability of having the same class variables in the same sub-function.

A PUBO_i is then characterized by 8 parameters: n (problem dimension), m (number of sub-functions), C (portfolio of sub-functions), p_i (probabilities of sub-function class), k (number of variable importance classes), n_i (number of variables in each importance class), d_i (degree of importance of each class), and α (co-appearance factor).

2.3 Local Optima Networks

The fitness landscape model [17] initially proposed in the biology field is nowadays regularly considered in the field of evolutionary computation to gain insights on the instance structure from the

perspective of neighborhood-based evolutionary algorithms. While fitness landscapes can be characterized through several properties and indicators [7, 8, 13] that provide valuable information, many of them fail to provide sufficient insights on the global structure of landscapes. A possibility to overcome this issue is to consider local optima networks (LONs) [10].

This coarse-grained model provides a compressed view of fitness landscapes allowing the visualization of some of its global characteristics, such as the local optima distribution, and provides a new set of metrics. LONs are networks in which nodes are local optima, and edges are transitions among these local optima. Therefore a given LON depends upon the neighborhood operator for its nodes and a perturbation operator for its links. In the following, we consider compressed monotonic LONs (CMLONs) [11], a LON variant in which edges correspond to non-deteriorating transitions only, and local optima plateaus are grouped in a single node. A plateau refers to a set of connected local optima with the same fitness value. This model is coarser than classical LONs and was proposed to better account for neutrality.

Formally, a CMLON is the directed graph (CL, CE) , where the compressed local optima CL are the nodes, and the edges CE are aggregated from the monotonic edge set E_m . There is a *monotonic edge* between two local l_1 and l_2 if one of them can be obtained after applying a random perturbation to the other one, followed by local search and if $l_1 \leq l_2$. Monotonic edges are weighted with estimated frequencies of transition. A *compressed local optimum* corresponds to a set of nodes with the same fitness value that are connected by monotonic edges.

A *monotonic sequence* is a path $MS = \{cl_1, \dots, cl_s\}$ of connected nodes ($cl_i \in CL$). By definition of the edges, $f(cl_i) \leq f(cl_{i-1})$. There is a natural end to every monotonic sequence, cl_s , when no improving transitions can be found. Then, cl_s is called a *sink* as it does not have outgoing edges. A *funnel* in the CMLON contains the aggregation of all monotonic sequences ending at the same sink.

3 GLOBAL LANDSCAPE STRUCTURE

3.1 Experimental Protocol

This section is focused on the global analysis of some PUBO_i landscapes through LONs and CMLONs metrics and visualization. LONs are sampled and constructed using an iterated local search (ILS) with a 1-flip neighborhood that alternates a best-improvement hill-climber with a perturbation phase that randomly changes the value of 10% of the solution bits. In the following we consider four PUBO_i instances with varying parameters (see table 1). PUBO_i 9009 and PUBO_i 9012 have no variable importance, and their difference lies in the sub-problems repartition. Instances PUBO_i 9010 and PUBO_i 9011 have two classes of importance and vary in their co-appearance factor (PUBO_i 9011 having the highest one). For each instance, we produce a LON and a CMLON by aggregating the unique nodes and edges encountered across 30 independent ILS runs.

3.2 Network Metrics

Table 2 shows metrics for the extracted LON models. For the standard LON model, we only report the number of nodes, the rest of the metrics refer to the CMLON model, as our goal is to study the landscapes' global structure. The provided metrics are:

- *nodes*: number of nodes (local optima) in the LON model.

Table 1: Instance parameters of PUBO_i. $n = 100, m = 500$

Id	(d_0, d_1)	α	(p_1, p_2, p_3, p_4)
9009	(1, 1)	1.0	(0.25, 0.25, 0.25, 0.25)
9010	(1, 4)	1.0	(0.25, 0.25, 0.25, 0.25)
9011	(1, 4)	1.15	(0.25, 0.25, 0.25, 0.25)
9012	(1, 1)	1.0	(0.10, 0.00, 0.00, 0.90)

- *comp. nodes*: number of compressed nodes.
- *sinks*: number of sub-optimal sinks.
- *bests*: number of nodes with the best fitness value.
- *best strength*: normalized incoming strength of the best nodes, where *strength* refers to the aggregated weight of incoming edges to the best nodes (normalized).
- *path to best*: average path length, in edges, from source nodes to best nodes.

Table 2 shows that the compressed monotonic (CMLON) model, conveys a coarser perspective of the landscapes' global structure. For all instances, there are more than one best node and multiple sinks. We can appreciate a multi-funnel structure where the number of sub-optimal funnels (sinks) generally outnumbers the number of optimal funnels (bests). 9011 contains the largest number of nodes and compressed nodes; it is also the easiest instance to solve as indicated by the high strength value of the best nodes. The strength of the best nodes indicates how much they "attract" the search process: a high strength represents an easy landscape. According to this metric, the hardest landscape studied is 9009. In terms of the average path to best solutions, 9011 is the instance requiring more search steps to hit a best solution, but there are two best large funnels, indicating that a best solution is reached often.

Table 2: Network metrics for all PUBO_i instances

inst.	nodes	comp. nodes	sinks	bests	best str.	path to best
9009	1169	203	7	2	0.38	6.93
9010	2877	194	3	2	0.57	6.25
9011	5975	297	2	2	0.9	11.18
9012	2047	223	9	4	0.48	7.65

3.3 Network Visualization

CMLONs visualization for each instance is provided in Figure 1. The CMLONs of the two instances with variable importance display larger best nodes (in red) and a higher number of global optima. Their global optima are often part of the same funnel. While large best nodes are also found in the 9012 CMLON, the high number of connected components ending up in a sink (in blue) wanes the likelihood to attain a global optimum. Indeed, the connected components are more likely to end in a sink than a global optimum on the CMLONs of instances without variable importance. For all instances, the CMLON has a global optimum located in its largest connected component, meaning that the largest funnels end in a global optimum. However, 9009 and 9010 also have a connected component close in size ending up in a sink. The proportion of sinks 9009 makes a global optimum harder to attain than on 9010.

The various observations highlight that instances without variable importance seem harder to solve considering the attainability of the global optima. All instances have many global optima, yet on landscapes without variable importance the multi-funnel structure,

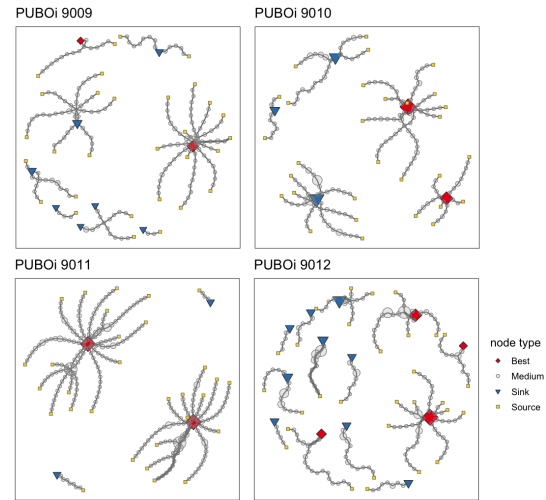


Figure 1: CMLONs sampled with an ILS for 4 PUBO_i instances with the legend provided on the right-hand side. incoming strength of global optima and ratio of sinks/best nodes have a negative impact on the attainability of global optima, as commonly observed in randomly generated academic instances. The sub-problem distribution has an effect on the attainability and neutrality levels, 9012 is mainly composed of a single sub-problem and has more neutrality and a better attainability than 9009.

The higher co-appearance factor seems to be an important parameter to achieve a global structure similar to that of real-world landscapes. Indeed, results on 9011 show its correlation with the size of global optima funnels and the best nodes/total nodes ratio. This instance also has by far the highest strength, as on CMLONs of real-world instances of the quadratic assignment problem [9].

4 PERFORMANCE ANALYSIS

4.1 Experimental Protocol

To conduct the performance analysis we consider the Deterministic Recombination and Iterated Local Search algorithm (DRILS), that combines a hill climber with both classic perturbation and a partition crossover [1]. We consider two DRILS variants that use a first improvement hill-climber and a 1-flip neighborhood, and varies in the perturbation operator: one uses random perturbation operator that flips $\kappa\%$ of the decision variables, and the other an operator based on the variable interaction graph (VIG) proposed in [15].

In the following, we perform a first set of experiments on the four previously analyzed instances 1, for which we aim to observe the possible differences in their solving. A second set of experiments is carried out on a set of larger instances ($n = 500, m = 5000$) to observe possible differences. For these instances, the sub-problems distribution remains always the same (25% probability for each) to focus on the variation of α . We consider instances without and with variable importance ($(d_0, d_1) = (1, 4)$), and with different co-appearance factor values $\alpha \in \{1.00, 1.05, 1.10, 1.15\}$. We use 20 instances for all possible combinations.

For both DRILS, we use a perturbation strength of $0.1 \times n$. We conducted 30 independent runs per instance and perturbation operator, with a stopping criterion of 5 seconds. Experiments were carried out on a Macbook with an M1 chip and 16Go of RAM. Data, and code are made available on: <https://gitlab.com/verel/pubo-importance-benchmark>.

4.2 Experimental analysis

Results on smaller instances are reported in table 3. Success rates show that instances 9011 and 9012 are always solved for both perturbation operators, while 9009 and 9010 are always solved with the vig perturbation operator. According to this and the duration required to achieve the best solution, the easiest instance to solve is 9011, followed by 9012. While the latter has no variable importance, its sub-problem distribution differs from other instances (only two different sub-problems).

Table 3: DRILS performance on smaller instances with a stopping criterion of 5s. Values in bold indicate statistical difference according to a Mann-Whitney test (p-value = 0.05), or a Chi2 test for the success rate.

Id	Avg. Fitness		Succ. rate		Time to best	
	rnd	vig	rnd	vig	rnd	vig
9009	-620.67	-622.00	0.833	1.000	2.670	1.673
9010	-481.60	-482.00	0.900	1.000	2.668	1.170
9011	-504.00	-504.00	1.000	1.000	0.005	0.172
9012	-475.00	-475.00	1.000	1.000	1.008	0.675

Figure 2 shows the performance comparison of DRILS according to the co-appearance parameter α of large PUBO_i instances, for the average best-found fitness, also averaged over the 20 independent instances for each value of α . The only varying PUBO_i parameter is the spreading of variables over sub-functions. Since PUBO_i are minimization problems, finding a lower fitness value suggests the constraints between variables makes it possible to attain solutions closer to the minimal bound. Consequently, instances with lower fitness found by DRILS algorithms can be interpreted as being easier to solve. For all large instances, the VIG perturbation outperforms the classical random perturbation. Above all, the performances of both DRILS improve when the co-appearance factor increases. The gap of performance between the two DRILS slightly wanes with the co-appearance factor α , though there is no change in relative performance between algorithms with α . As suggested by the previous LON analysis, instances from $\alpha = 1.0$ to $\alpha = 1.15$ seem to be easy to solve for DRILS algorithms. Notice that for the larger instances, the ratio between n and m is different from the one of smaller instances, and that the available computation time budget for the large instances is still limited to 5 seconds. Thus, the performance analysis could be contrasted with further and deeper investigations.

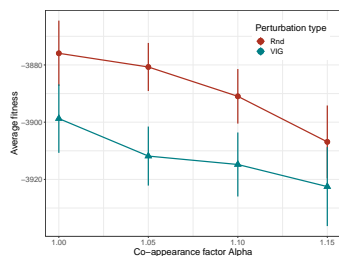


Figure 2: Average fitness (with confidence interval) on 20 instances of DRILS performances on instances with $n = 500$, and different co-appearance factor α .

5 CONCLUSION

In this work, we proposed a study of PUBO_i landscapes, a generator enabling to tune variable importance. By conducting a local optima network (LON) analysis, we found major differences in instances with and without variable importance, especially when the latter has a high co-appearance factor. PUBO_i instances with variable importance and a high co-appearance factor seem to be close to real-world landscapes in terms of global structure and LON metrics.

Then we conducted the first-ever performance analysis of PUBO_i landscapes using two DRILS variants. These first experiments show that for larger instances, the variable importance and co-appearance factor impact the performance of DRILS.

In future works, one should investigate the effect of various PUBO_i parameter values on the global structure of landscapes: the sub-problems distributions, levels of variable importance, and ratio between the number of variables and terms. Another goal would be to observe the impact of PUBO_i parameters on the performance of evolutionary algorithms.

REFERENCES

- [1] F. Chicano, D. Whitley, G. Ochoa, and R. Tinós. 2017. Optimizing one million variable NK landscapes by hybridizing deterministic recombination and local search. In *GECCO 2017*. 753–760. <https://doi.org/10.1145/3071178.3071285>
- [2] P. Date, R. M Patton, C. D Schuman, and T. E Potok. 2019. Efficiently embedding QUBO problems on adiabatic quantum computers. *Quantum Information Processing* 18, 4 (2019), 117. <https://doi.org/10.1007/s11128-019-2236-3>
- [3] A. Galluccio, M. Loebel, and J. Vondrák. 2001. Optimization via enumeration: a new algorithm for the max cut problem. *Mathematical Programming* 90, 2 (2001), 273–290. <https://doi.org/10.1007/s101070100217>
- [4] Fred Glover, Jin-Kao Hao, and Gary Kochenberger. 2011. Polynomial unconstrained binary optimisation—part 1. *International Journal of Metaheuristics* 1, 3 (2011), 232–256. <https://doi.org/10.1504/IJMEUR.2011.041196>
- [5] F. Hamze, D. C Jacob, A. J Ochoa, D. Perera, W. Wang, and H. G Katzgraber. 2018. From near to eternity: spin-glass planting, tiling puzzles, and constraint-satisfaction problems. *Physical Review E* 97, 4 (2018). <https://doi.org/10.1103/PhysRevE.97.043303>
- [6] Zhipeng Lü, Fred Glover, and Jin-Kao Hao. 2010. A hybrid metaheuristic approach to solving the UBQP problem. *European Journal of Operational Research* 207, 3 (2010), 1254–1262. <https://doi.org/10.1016/j.ejor.2010.06.039>
- [7] Katherine Mary Malan. 2021. A survey of advances in landscape analysis for optimisation. *Algorithms* 14, 2 (2021), 40. <https://doi.org/10.3390/a14020040>
- [8] Katherine M Malan and Andries P Engelbrecht. 2013. A survey of techniques for characterising fitness landscapes and some possible ways forward. *Information Sciences* 241 (2013), 148–163. <https://doi.org/10.1016/j.ins.2013.04.015>
- [9] G. Ochoa and S. Herrmann. 2018. Perturbation strength and the global structure of QAP fitness landscapes. In *PPSN XV, September, 2018, Part II 15*. Springer, 245–256. https://doi.org/10.1007/978-3-319-99259-4_20
- [10] G. Ochoa, M. Tomassini, S. Verel, and C. Darabos. 2008. A study of NK landscapes’ basins and local optima networks. In *Genetic and Evolutionary Computation Conference - GECCO 2008*. ACM, 555–562. <https://doi.org/10.1145/1389095.1389204>
- [11] G. Ochoa, N. Veerapen, F. Daolio, and M. Tomassini. 2017. Understanding Phase Transitions with Local Optima Networks: Number Partitioning as a Case Study. In *Evolutionary Computation in Combinatorial Optimization, (EVO-COP) (LNCS, Vol. 10197)*. Springer, 233–248. https://doi.org/10.1007/978-3-319-55453-2_16
- [12] D. Perera, I. Akpabio, F. Hamze, S. Mandra, N. Rose, M. Aramon, and H. G Katzgraber. 2020. Chook—A comprehensive suite for generating binary optimization problems with planted solutions. *arXiv preprint arXiv:2005.14344* (2020). <https://doi.org/10.48550/arXiv.2005.14344>
- [13] Erik Pitzer and Michael Affenzeller. 2012. A comprehensive survey on fitness landscape analysis. *Recent advances in intelligent engineering systems* (2012), 161–191. https://doi.org/10.1007/978-3-642-23229-9_8
- [14] S. Tari, S. Verel, and M. Omidvar. 2022. PUBO_i: A Tunable Benchmark with Variable Importance. In *Evolutionary Comp. in Combinatorial Optimization (Part of EvoStar)*. Springer, 175–190. https://doi.org/10.1007/978-3-031-04148-8_12
- [15] R. Tinós, M. W Przewozniczek, and D. Whitley. 2022. Iterated local search with perturbation based on variables interaction for pseudo-boolean optimization. In *GECCO 2022*. 296–304. <https://doi.org/10.1145/3512290.3528716>
- [16] J. L. Walsh. 1923. A Closed Set of Normal Orthogonal Functions. *American Journal of Mathematics* 45, 1 (1923), 5. <https://doi.org/10.2307/2387224>
- [17] S. Wright. 1932. The Roles of Mutation, Inbreeding, crossbreeding and Selection in Evolution. *XI International Congress of Genetics* 8 (1932), 209–222.