Coarse-Grained Barrier Trees of Fitness Landscapes

Sebastian Herrmann[1(✉)], Gabriela Ochoa[2], and Franz Rothlauf[1]

[1] Department of Information Systems and Business Administration, Johannes Gutenberg-Universität, Jakob-Welder-Weg 9, 55128 Mainz, Germany
{s.herrmann,rothlauf}@uni-mainz.de

[2] Department of Computing Science and Mathematics, University of Stirling, Stirling FK9 4LA, Scotland
goc@cs.stir.ac.uk

Abstract. Recent literature suggests that local optima in fitness landscapes are clustered, which offers an explanation of why perturbation-based metaheuristics often fail to find the global optimum: they become trapped in a sub-optimal cluster. We introduce a method to extract and visualize the global organization of these clusters in form of a barrier tree. Barrier trees have been used to visualize the barriers between local optima basins in fitness landscapes. Our method computes a more coarsely grained tree to reveal the barriers between clusters of local optima. The core element is a new variant of the flooding algorithm, applicable to local optima networks, a compressed representation of fitness landscapes. To identify the clusters, we apply a community detection algorithm. A sample of 200 NK fitness landscapes suggests that the depth of their coarse-grained barrier tree is related to their search difficulty.

Keywords: Fitness landscape analysis · Barrier tree · Disconnectivity graph · Local optima networks · Big valley · Search difficulty · NK-landscapes

1 Introduction

To overcome the problem of getting stuck in a local optimum, many metaheuristics based on local search apply a perturbation operator. The perturbation is supposed to "kick" an algorithm away from the current region of the search space. This principle is known as iterated local search (ILS) [1], e.g. as implemented in the Lin and Kernighan Heuristic [2,3]. The "big valley" hypothesis [4] states that the local optima in many fitness landscapes are not randomly distributed, but clustered and surrounding the global optimum. Consequently, one might assume that once a local optimum has been reached, ILS-based algorithms should easily find the global optimum after a limited number of perturbations. However, we know that this is by no means the case in practice. An approach

G. Ochoa—Acknowledges funding from the Leverhulme Trust, UK [award number RPG-2015-395].

J. Handl et al. (Eds.): PPSN XIV 2016, LNCS 9921, pp. 901–910, 2016.
DOI: 10.1007/978-3-319-45823-6_84

to explain this observation is given in the most recent literature [4–7]: instead of one big valley, fitness landscapes consist of multiple clusters (or funnels). The existence of such a structure offers a new explanation for the search difficulty of landscapes: since the connections between clusters are sparse, perturbation steps fail to escape from sub-optimal clusters to the cluster of the global optimum.

The objective of this paper is to complement the recent literature on the multi-cluster structure of landscapes with a new approach to study this structure, and to draw conclusions on search difficulty. A method that has been used to characterize the structure of fitness landscapes are barrier trees [8]. A barrier tree shows in a hierarchical structure how the local optima basins are connected in the landscape. The leaf nodes are the local optima and the branching nodes are the saddle points connecting the basins [9]. Due to the ability of ILS to easily move from local optimum to local optimum, we are primarily not interested in the barriers between their basins. The core issue for ILS is that local optima are clustered. Thus, we need to study which barriers exist between these clusters. The method we introduce here addresses this purpose. It allows us to compute a coarse-grained barrier tree and to characterize the landscape on the level of clusters. To reveal the clustering structure of landscapes, local optima networks (LONs) [10] have been used. A LON is a compressed representation of a fitness landscape. In a LON, each node is a local optimum, and the edges represent the transitions of an algorithm between the basins around the local optima. A problem with LONs is that it can be difficult to visualize their structure when they consist of a large number of nodes and edges. To identify clusters in fitness landscapes, statistical measures have been applied to LONs, e.g. counting the network graph's connected components [5] or community detection [7].

Our contribution is a modified version of the "flooding algorithm", which accepts as an input (i) a LON of a fitness landscape and (ii) a pre-computed clustering structure of the LON. The output is a coarse-grained picture of the landscape which retains the global structure and allows the eventual visualization of larger landscapes. We demonstrate our method with instances of the Kauffman NK model. For each instance, we computed the LON and the clusters. We obtained the clusters by community detection with the Markov cluster algorithm [11], as proposed in an earlier study [7]. We analyze the resulting barrier trees by visual inspection and a statistical approach. We provide an indication how the structure of the barrier tree is related to the search difficulty of a landscape.

The article is structured as follows: Sect. 2 introduces the concept of fitness landscapes for the study of problems and heuristic search. In Sect. 3, we explain how to construct a standard barrier tree for fitness landscape analysis. In order to construct a coarse-grained barrier tree (based on the local optima clusters), we need a method to identify the clustering structure. In Sect. 4, we introduce local optima networks as a compressed representation of fitness landscapes, and the Markov cluster algorithm to reveal the clustering structure of a fitness landscape. In Sect. 5, we present the algorithm to calculate the coarse-grained barrier tree of a fitness landscape. We visualize instances and examine the search difficulty. A brief summary and conclusions are in Sect. 6.

2 Fitness Landscapes

The concept Fitness Landscapes was introduced to study the reproductive success of genotypes in theoretical biology [12]. Fitness landscapes have been adopted in combinatorial optimization to study the structure of problems and the dynamics of heuristic search. A fitness landscape is defined as a triplet of the search space S, the fitness function f, and the neighborhood structure $N(S)$. The search space S contains all valid solutions. The fitness function $f : S \rightarrow \mathbb{R}_{\geq 0}$ assigns a fitness value to each $s \in S$ (we assume non-negative values and a maximization problem). The neighborhood function $N : S \rightarrow \mathcal{P}(S)$ assigns a set of neighbors $N(s)$ to every $s \in S$. Two solutions are neighbors if they are mutually reachable by one step of local search.

A local optimum is a solution that has a higher fitness than its neighbors [13]. A higher number of local optima (modality) leads to a landscape that is more "rugged", which increases the search difficulty for local search-based algorithms [14]. A local optimum is surrounded by a *basin of attraction*. The basin around an optimum is the set of solutions from which the optimum attracts a local search algorithm. We define a function for the basin around a local optimum lo as $B : lo \rightarrow \mathcal{P}(S \backslash LO)$. B assigns an element from the set of all subsets (power set $\mathcal{P}$) of solutions over the search space to each local optimum $lo \in LO$ (the set of all local optima).

The Kauffman NK model of landscapes [15] is frequently used for the study of fitness landscapes. The NK model is a combinatorial optimization problem from the class of pseudo-Boolean functions. An instance is defined by the two parameters N and K, where N is the number of binary variables. The size of the search space S is $|S| = 2^N$. K is the number of variables interacting with each other (epistasis). To instantiate the model, the co-variables are randomly selected. A higher value of K leads to a higher search difficulty [14]. The distance between two solutions $x, y \in S$ is the number of differing bits (Hamming-distance).

3 Barrier Trees of Fitness Landscapes

Barrier trees were introduced in computational chemistry to study the structure of potential energy landscapes [16,17], i.e. to examine the barriers that exist between the optima basins. Barrier trees are sometimes referred to as *disconnectivity graphs* [18,19]. Even though Barrier trees have been used to study heuristic search [8,9], the literature on this topic is rather sparse. To construct the barrier tree of a fitness landscape, a database of the local optima (we assume local maxima in this paper), and the transition states connecting at least two basins around different local optima, is required. The transition states are also called *s*addle points. In a 2-dimensional landscape, a saddle point is a local minimum. In a higher dimensional landscape, multiple of local minima, connecting two basins, may exist. In such a case, the saddle point is the local minimum with maximal fitness. Since the fitness of the saddle point is lower than the fitness of the two connected local optima, it can be interpreted as a barrier between them: to move from one of the local optima to the other, an algorithm has to accept a fitness deterioration down to the level of the local minimum. To visualize the barrier tree, local optima are identified with leaves, while the branching nodes represent saddle points separating groups of local optima.

A method to compute the barrier tree of a fitness landscape is the so-called "flooding algorithm" [9]. We think that a comprehensive understanding of this method is essential; hence we depict the mechanism in Fig. 1. For a maximization problem, the algorithm iterates over all solutions in the search space in a descending order (in terms of fitness): the landscape is "flooded". When a local optimum is found, a node is added to the barrier tree (steps 1 and 2). When a saddle point is found, a branching node is added to the tree, and edges are added to connect the saddle point to the adjacent local optima. From here, the saddle point now represents the basins of all adjacent local optima (step 3, the basins are merged by the flooding). This procedure is repeated until the last local optimum or saddle point has been found (step 4).

Since we are interested in the barriers that exist between the clusters of local optima in a landscape, we present a variant of the flooding algorithm suitable for this purpose in Sect. 5. Before, we need to explain how to characterize funnels in fitness landscapes. For this purpose, we introduce a special representation of fitness landscapes known as local optima networks (LONs) and a method using this representation to characterize funnels in the next Sect. 4.

4 Clusters of Local Optima in Fitness Landscapes

Local Optima Networks (LONs) are a novel approach to study the structure of fitness landscapes [10] and have recently been used to reveal the structure of multiple clusters [5–7, 20]. LONs were originally inspired by the study of energy landscapes [21]. A LON is a complex network in which the nodes represent the local optima in a landscape (and their basins, resp.). The edges reflect an algorithm's transition between the basins. The concept of LONs allows the study of fitness landscapes from a network perspective and has the potential to deepen our understanding of metaheuristics and problems.

A network is a graph $G = (V, E)$ with the set of vertices V and the set of edges E. In a LON, the vertex set V contains the local optima of the fitness landscape. There exists an edge between two local optima if their basins are in some way connected, leading to a potential transition between the two local optima. An escape edge [22] is defined by the distance function of the fitness landscape d (minimal number of moves between two solutions): there exists a directed edge e_{xy} from local optimum lo_x to lo_y if there is a solution s such that $d(s, lo_x) \leq D \wedge s \in B(lo_y)$. The weight w_{xy} of edge e_{xy} is the probability that a search algorithm can escape from the local optimum lo_x into the basin around lo_y. The constant $D > 0$ determines the maximum distance that an algorithm uses during a perturbation step.

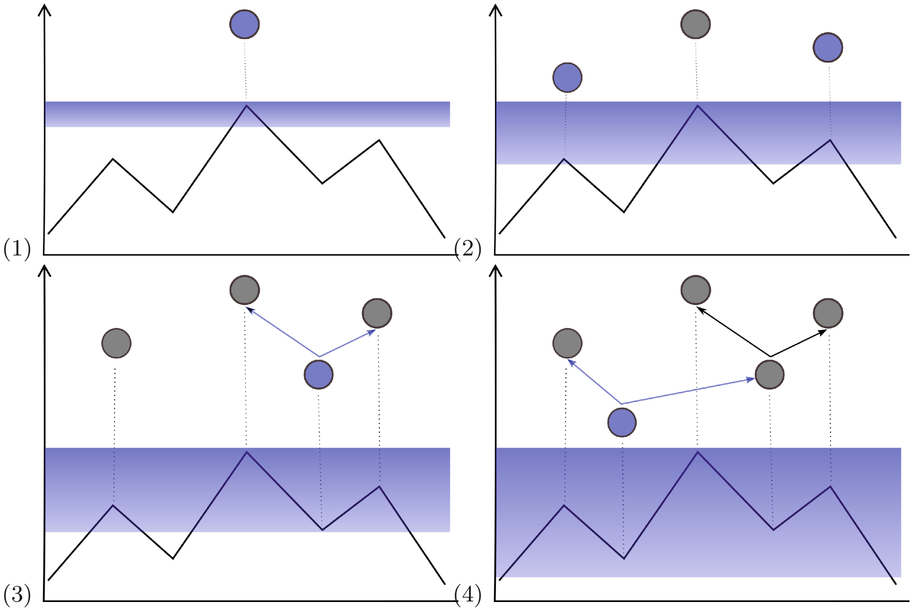

Fig. 1. Four steps of the flooding algorithm, creating the barrier tree of a fitness landscape. The vertical axis is the fitness, the horizontal axis is the landscape. Since we use a maximization problem, the space is "flooded" from the top to the bottom.

To reveal the clustering structure of fitness landscapes, we proposed to apply "community detection" to local optima networks [7]. Community detection is an exploratory variant of graph partitioning [23]. The objective of this method is to partition the network graph in a discipline-related, meaningful way. A very general definition of a community is a group of nodes that have more links among each other than to nodes in other communities. However, the definition of a community depends on the discipline applied and there exists a variety of algorithms that have been validated for different purposes [24,25].

Community detection in LONs has been done in earlier studies [26,27]. However, we [7] found that in particular, the Markov Cluster Algorithm (MCL, [11]) is an appropriate method of community detection to detect clusters in LONs and characterize the clustering structure of fitness landscapes. An explanation for this is that MCL is based on stochastic flows. LONs model the stochastic process of an algorithm in a fitness landscape. For this reason, the application of MCL matches the network model and produces meaningful results.

5 Coarse-Grained Barrier Trees of Fitness Landscapes

In order to escape from a cluster of local optima to another cluster, ILS needs to pass a barrier by a deterioration of the fitness. To visualize the structure of the barriers between the clusters in the landscape, we present a variant of the flooding algorithm [9] as introduced in Sect. 3 and Fig. 1. The pseudo code can be obtained from Algorithm 1. As an input, the algorithm accepts a LON and a

Algorithm 1. Flooding Algorithm for LONs (Maximization Problem)

Require: Local Optima Network $G = (V, E)$, Partition P over V (the cluster sets)
Let R be an empty set
for all $p \in P$ **do**
  Add the local optimum of p with max. fitness to R
end for{R contains one representing local optimum per cluster in P}
Let $T = (V_{Tree}, E_{Tree})$ be the empty Barrier Tree
Order V by f in descending order
for all $v \in V$ **do**
  if $v \in R$ **then**
    Add Node v to Barrier Tree V_{Tree}
  else
    $C = \{p \in P \mid \exists n \in p \mid ((v, n) \in E \vee (n, v) \in E)\}$
    {Select those partition sets (clusters) which contain a local optimum adjacent to v in the LON graph}
    if $|B| > 1$ **then** {v connects at least two clusters, i.e. v is a saddle point}
      Add Node v to Barrier Tree V_{Tree}
      for all $c \in C$ **do** {For each cluster set c connected to saddle point v}
        $r = c \wedge R$ {Choose node r representing connected cluster set c}
        Add Edge (v, r) to E_{Tree}
        Update P: Merge Partition set containing v and c
        Remove r from R {Flood the connected cluster}
      end for
    end if
  end if
end for
return T

partition of the LON's vertex set, i.e. a set with the clustering structure of the landscape. To obtain the clusters, we propose to apply the Markov cluster algorithm to the LON. As a first step, the algorithm selects the best local optimum for each cluster (set R). Then, the set of local optima nodes V is ordered by fitness in descending order. The algorithm iterates over each node. If the node is a representing node (in R), it is added to the barrier tree. Else, the algorithm determines the number of clusters adjacent to the current node in the LON. If the number is higher than one, the node is a saddle point and is also added to the tree. Then, the algorithm connects the saddle point to the nodes representing the adjacent clusters in the tree. From here, the saddle point represents all adjacent clusters ("flooding"): the clusters of the current and all the adjacent nodes are merged in the partition set, and the representers of the adjacent clusters are removed from R. This process is repeated until the whole LON is flooded (merged into one partition).

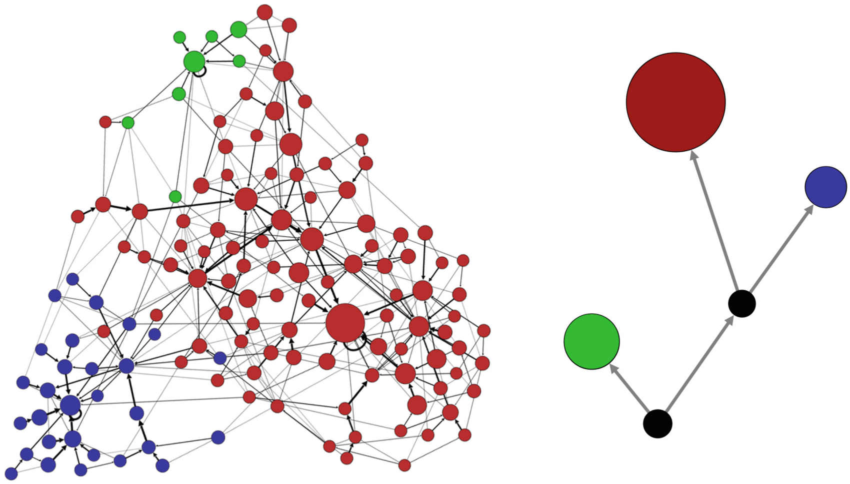

Fig. 2. Local optima network (left) and the coarse-grained barrier tree (right) of an NK landscape ($N = 20$, $K = 5$) with low search difficulty (success rate of ILS: 0.76). The color of the nodes represents the cluster (global optimum cluster is red in both graph types). In the tree, the branching nodes are black. In the local optima network, the size of the node represents the fitness, whereas the node size in the tree is the size of the cluster by the number of local optima. In the tree, the fitness is visualized by the node height (higher distance to the root means higher fitness). The layout of the local optima network is based on the ForceAtlas2 algorithm [28]. The local optima network shows only the best 20 % of nodes (all clusters still visible). (Color figure online)

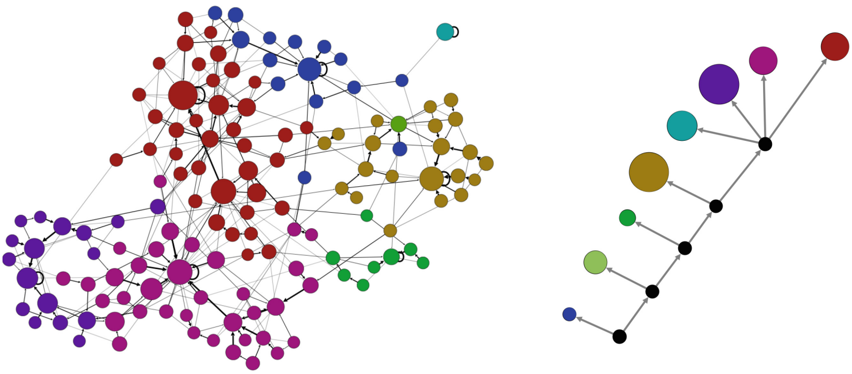

Fig. 3. LON and Coarse-Grained Barrier Tree of an NK landscape ($N = 20$, $K = 5$) with high search difficulty (success rate of ILS: 0.22). Please cf. Fig. 2 for further explanations.

To demonstrate our method, we selected an easy and a hard instance of the Kauffman NK model ($N = 20$, $K = 5$). To determine their difficulty, we performed 1000 independent runs of ILS per instance and measured the success rates (0.76 and 0.22). The ILS stopped after a limited number of fitness function evaluations ($^1/_{5th}$ of the search space), or when the global optimum was found. We extracted the LONs and computed the clusters in both LONs with MCL. We used the LONs and the clusters to construct the coarse-grained barrier trees with our variant of the flooding algorithm. Figures 2 and 3 plot the LON and the corresponding tree. Visual inspection of the LONs (left) confirms that the clustering as obtained by MCL is meaningful: nodes of the same color have a higher proximity to their own cluster than to those of a different cluster. Comparing both barrier trees (right), we observe a much deeper tree and thus a higher number of barriers in the case of the hard instance.

Even though a deeper study on the search difficulty is out of the scope of this paper, we conducted a first systematic approach towards this observation. We generated 200 instances of NK landscapes ($N = 20$, $K = 5$). We grouped the landscapes by the depth of the coarse-grained barrier tree and compared their difficulties for ILS. The results can be obtained from Fig. 4. For landscapes with a very short tree, we observe that the difficulty has a high variety, even though the median indicates a low difficulty (≈ 0.6). The median success rates get lower with a deeper tree, which means that their difficulty increases. This is not surprising: a deeper tree means that a traversal to the global optimum has—by average—a longer path. A search algorithm needs to pass more barriers then, and the difficulty is higher. This finding is consistent with the previous literature on regular barrier trees [9], however the observation that many landscapes with a low number of barriers can be difficult is counter-intuitive. We suggest that in these cases, additional factors, like the cluster size of the global optimum [7] need to be considered. We plan to conduct more research towards this direction.

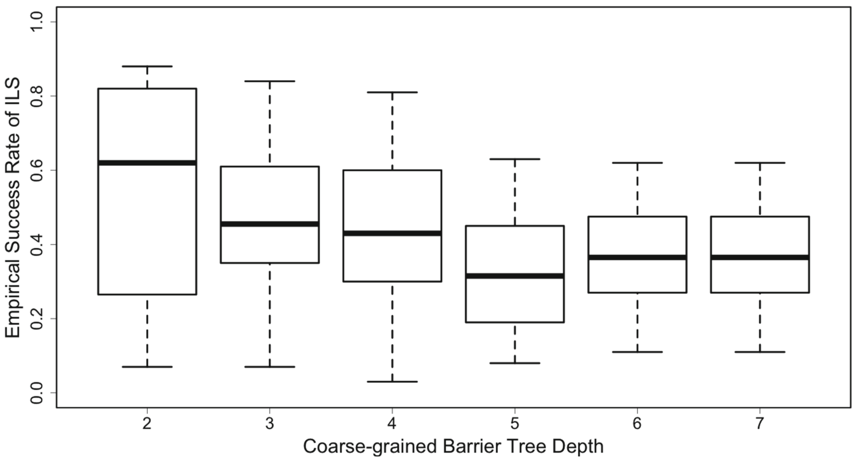

Fig. 4. Success Rate of ILS (difficulty) for different values of tree depth. The median success rate declines (search difficulty increases) with a higher depth of the tree.

6 Summary and Conclusion

As our main contribution, we presented a new method to visualize fitness landscapes and characterize them by the barriers between clusters of local optima. The existence of a multiple-cluster structure has recently emerged [6,7] as a refinement of the big valley hypothesis. We applied our method to a limited set of instances of the Kauffman NK model. Our results suggest that the tree depth might be related to the search difficulty of the landscapes for iterated local search. This is consistent with previous findings on difficulty in the literature [9]. A possible explanation is that the existence of barriers prevents iterated local search from escaping local optima clusters. This finding is rather preliminary and needs further investigation. Other structural properties of the landscapes must be taken into consideration, too. For further research, it would be interesting to see how the coarse-grained trees look for NK landscapes with higher levels of epistasis. It is also unclear whether or not there are differences between the NK model with random and adjacent co-variables. The adjacent NK model is often considered to be solvable with less effort. It would be worthwhile to examine if the tree depths are different between both models. We think that the method introduced here points to a new direction in studies of fitness landscapes.

References

1. Lourenço, H.R., Martin, O.C., Stützle, T.: Iterated local search. In: Handbook of Metaheuristics, pp. 320–353. Kluwer Academic Publishers, Boston (2003)
2. Applegate, D., Cook, W., Rohe, A.: Chained Lin-Kernighan for large traveling salesman problems. INFORMS J. Comput. **15**, 82–92 (2003)
3. Lin, S., Kernighan, B.W.: An effective heuristic algorithm for the traveling-salesman problem. Oper. Res. **21**, 498–516 (1973)
4. Hains, D.R., Whitley, D.L., Howe, A.E.: Revisiting the big valley search space structure in the TSP. J. Oper. Res. Soc. **62**, 305–312 (2011)
5. Ochoa, G., Veerapen, N., Whitley, D., Burke, E.K.: The multi-funnel structure of TSP fitness landscapes: a visual exploration. In: Bonnevay, S., Legrand, P., Monmarché, N., Lutton, E., Schoenauer, M. (eds.) EA 2015. LNCS, vol. 9554, pp. 1–13. Springer, Heidelberg (2016). doi:10.1007/978-3-319-31471-6_1
6. Ochoa, G., Veerapen, N.: Deconstructing the big valley search space hypothesis. In: Chicano, F., Hu, B., García-Sánchez, P. (eds.) EvoCOP 2016. LNCS, vol. 9595, pp. 58–73. Springer, Heidelberg (2016). doi:10.1007/978-3-319-30698-8_5
7. Herrmann, S., Ochoa, G., Rothlauf, F.: Communities of local optima as funnels in fitness landscapes. In: Proceedings of 2016 Genetic and Evolutionary Computation Conference - GECCO 2016 (2016)
8. Hallam, J., Prügel-Bennett, A.: Large barrier trees for studying search. IEEE Trans. Evol. Comput. **9**, 385–397 (2005)
9. van Stein, B., Emmerich, M., Yang, Z.: Fitness landscape analysis of nk landscapes and vehicle routing problems by expanded barrier trees. In: Emmerich, M., et al. (eds.) Evolutionary Computation IV. AISC, vol. 227, pp. 75–89. Springer, Heidelberg (2013)

10. Ochoa, G., Tomassini, M., Vérel, S., Darabos, C.: A study of NK landscapes' basins and local optima networks. In: Proceedings of 10th Annual Conference on Genetic and Evolutionary Computation - GECCO 2008, p. 555. ACM Press, New York (2008)
11. van Dongen, S.: Graph clustering by flow simulation. Ph.D. thesis, Utrecht University (2001)
12. Wright, S.: The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In: Proceedings of 6th International Congress of Genetics, pp. 356–366 (1932)
13. Glover, F.: Future paths for integer programming and links to artificial intelligence. Comput. Oper. Res. **13**, 533–549 (1986)
14. Weinberger, E.: Correlated and uncorrelated fitness landscapes and how to tell the difference. Biol. Cybern. **336**, 325–336 (1990)
15. Kauffman, S.A., Weinberger, E.D.: The NK model of rugged fitness landscapes and its application to maturation of the immune response. J. Theor. Biol. **141**, 211–245 (1989)
16. Becker, O.M., Karplus, M.: The topology of multidimensional potential energy surfaces: theory and application to peptide structure and kinetics. J. Chem. Phys. **106**, 1495 (1997)
17. Flamm, C., Hofacker, I.L., Stadler, P.F., Wolfinger, M.T.: Barrier trees of degenerate landscapes. Zeitschrift füer Physikalische Chemie **216**, 155 (2002)
18. Doye, J.P.K., Miller, M.A., Wales, D.J.: Evolution of the potential energy surface with size for Lennard-Jones clusters. J. Chem. Phys. **111**, 8417 (1999)
19. Doye, J.P.K., Miller, M.A., Wales, D.J.: The double-funnel energy landscape of the 38-atom Lennard-Jones cluster. J. Chem. Phys. **110**, 6896 (1999)
20. Ochoa, G., Veerapen, N.: Additional dimensions to the study of funnels in combinatorial landscapes. In: Proceedings of 2016 Genetic and Evolutionary Computation Conference - GECCO 2016 (2016)
21. Stillinger, F.H.: A topographic view of supercooled liquids and glass formation. Science **267**, 1935–1939 (1995)
22. Vérel, S., Daolio, F., Ochoa, G., Tomassini, M.: Local optima networks with escape edges. In: Hao, J.-K., Legrand, P., Collet, P., Monmarché, N., Lutton, E., Schoenauer, M. (eds.) EA 2011. LNCS, vol. 7401, pp. 49–60. Springer, Heidelberg (2012)
23. Talbi, E., Bessière, P.: A parallel genetic algorithm for the graph partitioning problem. In: Proceedings of 5th International Conference on Supercomputing - ICS 1991, pp. 312–320. ACM Press, New York (1991)
24. Fortunato, S.: Community detection in graphs. Phys. Rep. **486**, 75–174 (2010)
25. Porter, M.A., Onnela, J.P., Mucha, P.J.: Communities in networks. Not. AMS **486**, 1082–1097 (2009)
26. Daolio, F., Tomassini, M., Vérel, S., Ochoa, G.: Communities of minima in local optima networks of combinatorial spaces. Phys. A: Stat. Mech. Appl. **390**, 1684–1694 (2011)
27. Iclanzan, D., Daolio, F., Tomassini, M.: Data-driven local optima network characterization of QAPLIB instances. In: Proceedings of 2014 Conference on Genetic and Evolutionary Computation - GECCO 2014, pp. 453–460. ACM Press, New York (2014)
28. Jacomy, M., Venturini, T., Heymann, S., Bastian, M.: ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PLoS ONE **9**, e98679 (2014)