Gutenberg School of Management and Economics

& Research Unit "Interdisciplinary Public Policy"

Discussion Paper Series

Communities of Local Optima as Funnels in Fitness Landscapes

Sebastian Herrmann, Gabriela Ochoa and Franz Rothlauf

Discussion paper number 1609

Contact details:

Sebastian Herrmann
Department of Information Systems and Business Administration
Johannes Gutenberg University Mainz
Jakob-Welder-Weg 9
55128 Mainz, Germany
Email: s.herrmann@uni-mainz.de

Gabriela Ochoa
Department of Computing Science and Mathematics
University of Stirling
Stirling, FK9 4LA, Scotland
Email: goc@cs.stir.ac.uk

Franz Rothlauf
Department of Information Systems and Business Administration
Johannes Gutenberg University Mainz
Jakob-Welder-Weg 9
55128 Mainz, Germany
Email: rothlauf@uni-mainz.de

Communities of Local Optima as Funnels in Fitness Landscapes

Sebastian Herrmann
Department of Information
Systems and Business
Administration
University of Mainz, Germany
s.herrmann@uni-
mainz.de

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

Franz Rothlauf
Department of Information
Systems and Business
Administration
University of Mainz, Germany
rothlauf@uni-mainz.de

ABSTRACT

We conduct an analysis of local optima networks extracted from fitness landscapes of the Kauffman NK model under iterated local search. Applying the Markov Cluster Algorithm for community detection to the local optima networks, we find that the landscapes consist of multiple clusters. This result complements recent findings in the literature that landscapes often decompose into multiple funnels, which increases their difficulty for iterated local search. Our results suggest that the number of clusters as well as the size of the cluster in which the global optimum is located are correlated to the search difficulty of landscapes. We conclude that clusters found by community detection in local optima networks offer a new way to characterize the multi-funnel structure of fitness landscapes.

Keywords

Fitness landscape analysis; search difficulty; local optima networks; NK-landscapes.

1. INTRODUCTION

The analysis of fitness landscapes reveals that local optima are often not randomly distributed in the search space, but instead they are clustered in a "central massif" or "big valley". This so-called big valley hypothesis holds for a variety of optimization problems including the traveling salesman problem (TSP) [4]. Recent studies [10, 24, 23] extend the big valley hypothesis as they find that there is a structure of multiple funnels (instead of one single cluster) in the fitness landscape, which leads to a higher search difficulty for algorithms based on the principle of iterated local search (ILS), a search strategy that combines local search with perturbation steps [17]. Such global multi-funnel structures have been observed before in continuous optimization [16, 18, 14], however a more detailed characterization of funnels in combinatorial spaces like is still lacking.

Ochoa et al. [24] propose to characterize funnels in combinatorial spaces using local optima networks (LONs, [22]). The idea of LONs was inspired by the study of energy landscapes [28], and it was found that energy landscapes often have a structure of multiple funnels as well [19]. A local optima network is a network representation of a fitness landscape. In [24] the authors examined the LONs extracted from several TSP instances[1] under the *Chained Lin & Kernighan Heuristic* [2, 15], which is an ILS approach. The extracted LONs consisted of multiple components, and they conjectured that the absence of ties between these components could be an explanation of why ILS often fails to find the global optimum: since there is no path frequently connecting the components, the algorithm may get trapped in one of them. Consequently, components in LONs could offer a way to characterize funnels. This study considered large search spaces of the TSP instances, in consequence the LONs were collected by a sampling procedure. Thus, the presence of multiple components could be a consequence of the sampling. Furthermore, the notion of components conflicts with the formal definition of a fitness landscape, which usually consists of a single component.

We argue that LONs are a promising approach for a deeper study on the extended big valley hypothesis, and it would be worthwhile to examine if the existence of funnels can be shown by a state-of-the-art method from the portfolio of complex network analysis, i.e. "community detection" [7]. Furthermore, it would be interesting to examine systematically how the presence of funnels is related to search difficulty.

This paper explores whether a multi-funnel structure exists for landscapes from the Kauffman NK model [13] under iterated local search. The NK model is a class of pseudo-Boolean functions that have been used frequently in studies on fitness landscapes and search heuristics performance. When studying landscapes from the NK model, we are able to generate a large number of instances, and to limit the size of the search space. As a result, the computational effort for extracting local optima networks can be adjusted. Our study applies a "community detection" approach called the Markov Cluster Algorithm [31] to identify the funnels in LONs or landscapes, resp. Community detection has so far only been applied once on LONs [6], but the implications for heuristic search are yet unclear.

The article is structured as follows: section 2, describes

---

[1] http://comopt.ifi.uni-heidelberg.de/software/TSPLIB95/

the concept of fitness landscapes and local optima Networks with escape edges. Section 3, summarizes iterated local search. Section 4, describes our experimental setup. Our results are presented and discussed in section 5. A brief summary and our conclusions are in section 6.

2. FITNESS LANDSCAPES & LOCAL OPTIMA NETWORKS

Fitness landscapes are a concept that originated from theoretical biology [34]. In combinatorial optimization, the concept of fitness landscapes can be used to study the structure of problems as well as 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 solution candidates. The fitness function $f : S \to \mathbb{R}_{\geq 0}$ assigns a fitness value[2] to each $s \in S$. The neighborhood function $N : S \to \mathcal{P}(S)$ assigns a set of neighbors $N(s)$ to every $s \in S$. Usually, the neighbors are the solutions that can be reached by a local search step.

Local search is a concept that iteratively tries to improve a solution by applying small changes (in terms of the distance function). A simple implementation of local search is the best-improvement hill climber (algorithm 1). The algorithm usually starts with a random solution. It scans the neighborhood of the current solution and selects the best neighbor with a superior fitness as the next solution. This procedure is repeated until no better neighbor is found. Then, the algorithm has reached a local optimum and terminates.

A *Local optimum* is a solution that has a higher fitness than its neighbors. Local optima can not be overcome by a search method moving from a solution to one of its neighbors and accepting only better solutions [9]. A higher number of local optima leads to a landscape that is more "rugged", which generally indicates a higher search difficulty for local search [33].

A local optimum is surrounded by a *basin of attraction*, i.e. the set of solution candidates from which the hill climbing algorithm converges to the focal optimum. The basin around a local optimum lo is defined as a function

$$B : lo \to \mathcal{P}(S \backslash LO) \qquad (1)$$

which 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).

A *local optima network* (LON) [22] is a representation of a fitness landscape that allows the application of the complex-network analysis framework. Complex networks have been used to study the structure and dynamics of systems that consist of numerous entities which are in some way connected [1, 3]. Studies on the dynamics in networks include the influence of nodes (centrality) as well as information flow and diffusion [5, 30]. LONs are a novel way to examine the trajectory of algorithms in fitness landscapes.

A network is a graph $G = (V, E)$ with the set of vertices V and the set of edges E. For a LON, the vertex set V represents all local optima of the fitness landscape. An edge exists between two nodes (local optima), if there is a potential transition between the two local optima. The edges are directed and weighted. Edge weights $w_{x,y}$ represent the

---

[2]In this paper, we assume that the fitness function returns non-negative values.

Algorithm 1 Best-improvement Hill Climbing (hillClimb)

Require: Solution space S,
    Fitness function $f(S)$,
    Neighborhood function $N(S)$,
    Initial solution s_0
1: i $\leftarrow 0$
2: **repeat**
3:     choose $s_{i+1} \in N(s_i)$ s.t. $f(s_{i+1}) = max_{x \in N(s_i)}(f(x))$

4:     **if** $f(s_i) < f(s_{i+1})$ **then**
5:         $s_i \leftarrow s_{i+1}$
6:     **end if**
7:     $i \leftarrow i + 1$
8: **until** s_i is local optimum:
    $\{s \in N(s_i) \,|\, f(s) \geq f(s_i)\} = \{\}$
9: **return** s_i

---

probability that a search algorithm can move from local optimum lo_x to a solution in the basin around lo_y, assuming that the current state is lo_x. Vérel et al.[32] introduced the concept of escape edges, which are defined according to the distance function of the fitness landscape d (minimal number of moves between two solutions). An escape edge is defined as follows: there exists a directed edge e_{xy} (escape edge) pointing from local optimum lo_x to lo_y if there is a solution s such that

$$d(s, lo_x) \leq D \wedge s \in B(lo_y). \qquad (2)$$

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 is allowed for the escape. A LON with escape edges is a model describing the stochastic process of iterated local search (ILS) in a fitness landscape [12].

3. ITERATED LOCAL SEARCH

Iterated local search (ILS) combines the concept of intensification by local search with diversification by a number of perturbation steps. During intensification, heuristics focus on promising areas of the search space, whereas during diversification, new areas are explored [26]. Algorithm 2 describes the functionality of ILS. Usually, ILS starts with a randomly selected solution s_0 from the search space S. Then, the algorithm performs a hill climbing procedure (algorithm 1).

Hill climbing stops when it reaches a local optimum s^*, i.e no further improvement is possible. Then, ILS performs a diversification step by applying a limited perturbation to the local optimum, resulting in s'. As a next step, hill climbing is again applied starting with s', until the next local optimum $s^{*'}$ is reached. If the new local optimum $s^{*'}$ is different from the previous s^* and has higher fitness, the algorithm has "escaped" to a new local optimum, and the change is accepted. Otherwise, another perturbation is applied to s^*. This procedure is repeated until a termination condition is met, e.g. a fixed number of perturbation steps without any further improvement.

4. EXPERIMENTAL SETUP

For our experiments, we calculated the local optima networks for 300 instances of the Kauffman NK model [13]. The

Algorithm 2 Iterated Local Search (ILS)

Require: Solution space S,
    Fitness function $f(S)$,
    Neighborhood function $N(S)$,
    Stopping threshold t
1: Choose initial random solution $s_0 \in S$
2: $s^* \leftarrow \text{hillClimb}(s_0)$
3: $i \leftarrow 0$
4: **repeat**
5:    $s' \leftarrow \text{perturbation}(s^*)$
6:    $s*' \leftarrow \text{hillClimb}(s')$
7:    **if** $(s^{*'} \neq s^*) \land (f(s^{*'}) > f(s^*))$ **then**
8:       $s^* \leftarrow s^{*'}$
9:       $i \leftarrow 0$
10:    **end if**
11:    $i \leftarrow i + 1$
12: **until** $i \geq t$
13: **return** s^*

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 and K is the number of variables interacting with each other. The size of the search space S is $|S| = 2^N$. The fitness function

$$f_{NK} : [0,1]^N \rightarrow [0,1] \tag{3}$$

assigns a score to every combination of bits. The fitness $f_{NK}(s)$ of a solution s is the average of the values of N sub-functions (one for each bit). Each sub-function f_i assigns a fitness contribution for each bit i, depending on the value of bit i and K other bits that were randomly selected before instantiation:

$$f_i : [0,1]^{K+1} \rightarrow [0,1]. \tag{4}$$

The parameter K determines the number of co-variables per decision variable (epistasis). All values of the fitness function f_{NK} are normalized to values between 0 and 1, with $f_{NK}(s_{opt}) = 1$ as the fitness of the global optimum s_{opt}. In general, a higher value leads to a higher search difficulty [33]. The distance between two solutions $x, y \in S$ is calculated by the Hamming-distance $d(x,y) = \sum_{i=0}^{n} |x_i - y_i|$, i.e., the number of bits that are set to different values when comparing two solutions.

We randomly generated 300 NK fitness landscapes with $N = 20$ decision variables and different values of $K \in \{5, 10, 15\}$. Thus, we have 100 problems instances each for three levels of epistasis K. The size N of our problem instances is relatively low, since the computational effort for the experiments grows exponentially by the problem size N (especially calculating the LON is time-consuming). For each instance, we extracted the local optima network and applied the Markov Cluster Algorithm (MCL) on the networks to detect funnels in the landscapes.

Furthermore, we applied ILS to each problem instance and calculated the percentage of runs that are able to find the optimal solution. Results are averaged for 100 independent ILS runs for each problem instance. For the hill climbing steps procedure in ILS, we assumed that two solutions x, y are neighbors if their Hamming distance is equal to one ($d_{\max} = 1$). Thus, a local search step flips exactly one bit of the current solution. For the perturbation operator in ILS, we flip two random bits in one step. When extracting the LONs, we set the parameter D for the maximum escape distance to $D = 2$.

We calculated the following measures for each fitness landscape or LON, respectively:
- $\#lo$: the number of local optima, i.e. the number of nodes in the corresonding LON,
- $\#c$: the number of clusters that are found by MCL,
- $\#br$: the number of bridges (nodes that are in-between two clusters) that are found by MCL,
- fo: the size of the global optimum's cluster over the total number of local optima,
- mod and mod_w: two measurements of modularity to assess the quality of the clustering as proposed by the MCL algorithm (the concept of modularity will be introduced in section 5.3),
- $SRate_{ILS}$: the success rate of ILS to find the global optimum (averaged over 100 runs).

We used Java to generate the NK landscapes, to extract the LONs, and to apply ILS. We ran the experiments on a cluster using 64 cores with 256 GB of RAM per node. The running time per fitness landscape was approx. 20 minutes in the case of low epistasis, and 2 hours in the case of high epistasis. We implemented the Markov Cluster Algorithm using Python. Statistical analysis was done with R, visualizations of the graphs with *Gephi*.

5. COMMUNITY DETECTION ANALYSIS

5.1 Markov Cluster Algorithm

Community detection is method of graph partitioning. The objective of a graph partitioning problem is to search for a partition of a graph's nodes which optimizes a given cost function. A typical cost function is the number of links that connect between the partitions. Mostly, there are also several constraints, e.g. limits for the allowed number of partitions, nodes per group, etc. Many graph partitioning problems are NP-hard [29].

Community detection is a rather exploratory method in the sense that there are no pre-formulated constraints to the problem of choosing a partition of a graph (or network, respectively). Instead, a community detection algorithm is free in determining the number of communities or the number of nodes per community. 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 [7, 25].

To select an algorithm for community detection in LONs, we took into consideration that a LON represents the stochastic process of an algorithm in the fitness landscape. An algorithm for detecting communities in graphs of stochastic flows is the Markov Cluster Algorithm (MCL, [31]). MCL has been successfully used in various domains, e.g. protein folding networks [27], and also in a study on local optima networks of the Quadratic Assignment Problem [6] to identify clusters of local optima.

A description of MCL is given in algorithm 3. Let G be the graph of a LON. E is the adjacency matrix containing the weights of the directed edges in G. Since the edge weights

Algorithm 3 Markov Cluster Algorithm (MCL)

Require: Graph G,
  Power Parameter p,
  Inflation Parameter r
1: $E = AdjacencyMatrix(G)$
2: **repeat**
3:   {Expand Matrix}
4:   $E = E^p$
5:   **for all** $A_i \in E$ **do**
6:     {Inflate Columns}
7:     $E_i = A_i^r$
8:     $E_i = normalizeVector(E_i)$
9:   **end for**
10: **until** E has converged
11: **return** E

---

are the non-negative probabilities to move from a given local optimum into the basin around another local optimum, the probabilities of all columns in E sum up to 1. Thus, E is a stochastic matrix, also known as a transition matrix in a discrete-time Markov chain.

To identify communities in G, MCL applies two mechanisms: expansion and inflation. The expansion operator raises the adjacency matrix E to the non-negative power p. Expansion ensures that different regions of the graph are connected. The second mechanism is the inflation operator. Inflation raises each column E_i from the adjacency matrix E to a non-negative power r, and then re-normalizes the column. The re-normalization ensures that each column again sums up to 1, which is a constraint for a stochastic matrix of a Markovian process. As a results, inflation increases the weights of heavy-weighted edges, whereas the weights of low-weighted edges are reduced. Both mechanisms are repeated until the algorithm converges, i.e. the transition matrix E reaches a steady state. For our NK landscapes with $N = 20$, this state is usually reached after ≈ 30 iterations.

Applying Markov Clustering on a LON of an NK landscape results in a new graph (see figure 2 for an example), where the different clusters can be identified. As a result of the clustering algorithm, we obtain several unconnected subgraphs, where each subgraph is a star: in each star, there is an "attractor" in its center and a periphery around it. Each star (sub-graph) represents a community (or cluster) of the original graph (figure 1), and thus each node in the original graph belongs to one of these partitions, as indicated by the colors. Sometimes, we find nodes that belong to two clusters (for example, the purple node in figure 2). In our analysis, we call such nodes *bridges*.

5.2 Community Structure of the LONs

Figure 1 plots an example of a LON of an NK landscape with low epistasis ($K = 5$). An edge between two nodes indicates the existence of an escape edge. All nodes that are assigned to the same cluster by MCL have the same color. The size of a node indicates the fitness (larger size means better fitness). We only plot the best 10% of the nodes. In the plot, the nodes have been positioned by a force-directed layout ("ForceAtlas2"). This algorithm arranges the nodes in an aesthetically pleasing way by simulating that the edges between nodes are springs, and then tries to minimize the tension of the springs as well as the number of intersections.

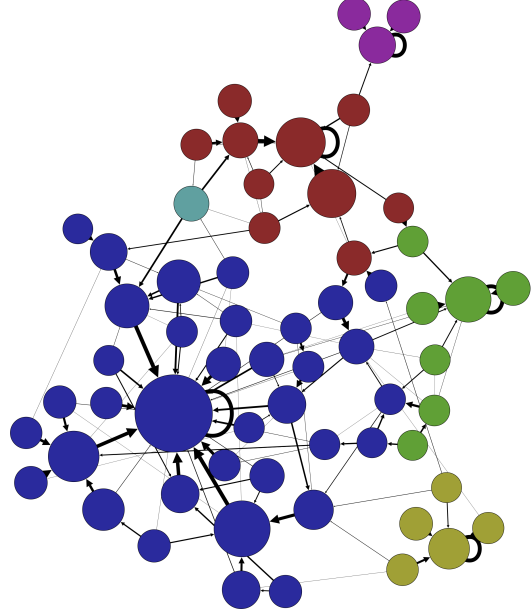Applying the MCL algorithm to this LON yields the net-

Figure 1: Example of a Local Optima Network of an NK landscape with $N = 20$, $K = 5$. Each node is a local optimum. The color of the nodes represents the cluster assignment by MCL. The size of a node indicates its fitness. The edge thickness indicates the edge weight, which is the probability to move from the outbound local optimum into the basin around the inbound local optimum. To highlight the structure of the network, we only plot the best 10% of nodes (according to fitness).
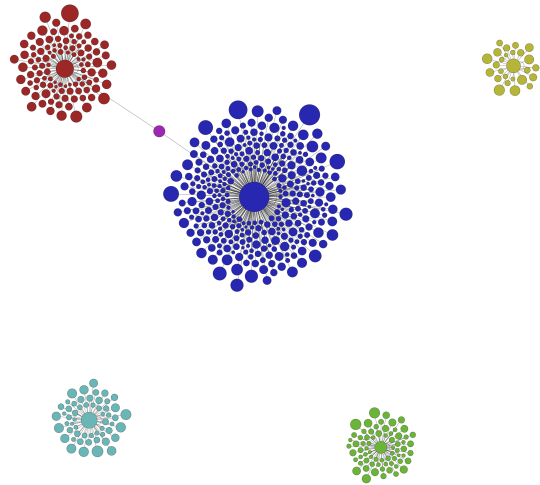

Figure 2: The output of the Markov Cluster Process applied to the LON from figure 1. As before, each node is a local optimum and the node size indicates fitness. Each subgraph represents a community, and nodes in the centers are their attractors. There is a bridge node connecting the blue and the red cluster.

K	$\overline{\#lo}$	$\overline{\#c}$	$\overline{\#br}$	$\overline{fo}$	$\overline{SRate_{ILS}}$
5	546.9	5.285	0.25	0.412	0.407
10	3384.5	19.985	2.303	0.147	0.142
15	7957	42.101	7.621	0.074	0.069

Table 1: Characteristics of networks obtained by Markov Clustering: average number $\overline{\#lo}$ of local optima, clusters $\overline{\#c}$, and bridges $\overline{\#br}$, average size $\overline{fo}$ of global optimum's cluster over the total number of local optima, and average success rate $\overline{SRate_{ILS}}$ of ILS

K	$\overline{Q}$	$\overline{Q_w}$
5	0.3721	0.4789
10	0.2629	0.3966
15	0.1947	0.3019

Table 2: Average modularity (with and without considering edge weights) for different values of K achieved by the Markov Cluster Algorithm for Community Detection.

work plotted in Figure 2: MCL iteratively increases the heavy weights and decreases the low weights until a new graph, consisting of several unconnected stars emerges. In the figure, we plot all nodes (and not only the best 10%).

The nodes in both figures have been colored according to the clusters that result from applying MCL on this instance. In Figure 1, we see that nodes that are identified to be in the same cluster are not randomly distributed across the graph but clustered. The nodes that are closely positioned to each other have the same color in most cases and thus belong to the same community. Even though we should be careful about intuitive conclusions from a visual inspection, the detection of the communities as proposed by MCL matches the visual topological structure of the LON.

As a next step, we study the characteristics of the networks obtained by Markov Clustering. In particular, we examine how the characteristics of LONs depend on the epistasis K. Table 1 lists the results. As expected, the average number of local optima increases with K. This explains the lower value of modularity with growing K: a higher number of nodes makes it more difficult to find a good partition of a graph. The total number of clusters also grows with K, and so does the average number of bridges. The fraction of the global optimum's cluster becomes lower for higher K. This effect also holds for the empirical success rate of ILS, which is a measurement for the search difficulty of the landscapes.

An additional finding we made is on the characteristics of the bridge nodes. Studying our data, we found that the bridges are among those nodes that have the highest closeness centrality in the LONs. Closeness centrality takes into account the geodesic (shortest-path) distances between all the nodes. The shorter a node's paths to all other nodes are, the higher is its closeness centrality [8]. An interpretation of this is that these nodes are by average "close" to the other nodes; they connect different areas of the fitness landscape. It would be worthwhile to study in future research whether typical characteristics of these nodes can be identified. This possibility might offer new opportunities for better search operators.

5.3 Quality of Community Structure

We want to quantify the quality of the community structure, i.e. the partition of the network. A common approach to quantify the strength of a community structure is the modularity Q as proposed by Newman and Girvan [21]. Given a certain partition of a graph, the modularity of this partition is "the number of edges falling within groups minus the expected number in an equivalent network with edges placed at random" [20]. Thus, $Q = 0$ indicates a partition

of the network by which the number of within-community edges is not better than random, whereas $Q = 1$ is a perfect partition of the network. In practice, values between $[0.3, 0.7]$ indicate a strong community structure [21].

The modularity of a graph can be calculated by either ignoring or considering the edge weights w_{ij}. We calculated both variants of modularity, where Q ignores the edge weights and Q_w considers the edge weights when calculating the modularity. Table 2 presents the results for the community structures revealed by MCL in our LONs. When ignoring weights, we observe a value of Q between 0.37 and 0.19. Taking the edge weights into account, we obtain higher values for Q_w between 0.48 and 0.30. In general, the modularity decreases with increasing epistatis K (see below). A possible explanation for the difference between Q and Q_w is that in LONs, the edge weights represent the transition probabilities between the local optima. Due to a non-linear distribution of these weights, the calculation of Q is biased towards lower values. In general, considering edge weights for the modularity returns more accurate results; values of $Q_w > 0.3$ indicate that the partition of the networks as proposed by MCL is satisfactory and we obtain meaningful clusters [21].

We conclude from the high quality of the community structure that the presence of several clusters in the LONs can be confirmed. Assuming that these cluster (communities) are an alternative way of characterizing "funnels", this finding underpins the hypothesis that there is a structure of multiple funnels in fitness landscapes of the NK model under ILS.

5.4 Community Structure and Search Difficulty

Finally, we examine the relationship between the community structure and search difficulty. Thus, we calculate the Pearson correlations between the number of clusters as well as the size of the cluster containing the global optimum and the ILS success rate. For the number of clusters, we observe a medium correlation ($R^2 \approx 0.46$, figure 3). Thus, the presence of multiple clusters has an effect on search difficulty, even though the correlation is low.

Figure 4 plots the correlation between the relative size fo of the cluster containing the global optimum and ILS success rate. We find a strong correlation between fo and search difficulty ($R^2 \approx 0.94$). Thus, the size of the cluster containing the global optimum is a strong predictor for the performance of ILS. A possible explanation for this correlation is that ILS can more easily find the global optimum if it is surrounded by many other local optima. In contrast, if the global optimum's cluster is very small, it is very likely that ILS gets stuck in another cluster and a high number of perturbations are necessary to reach the cluster of the global

$R^2 = 0.4559238$ (left figure)
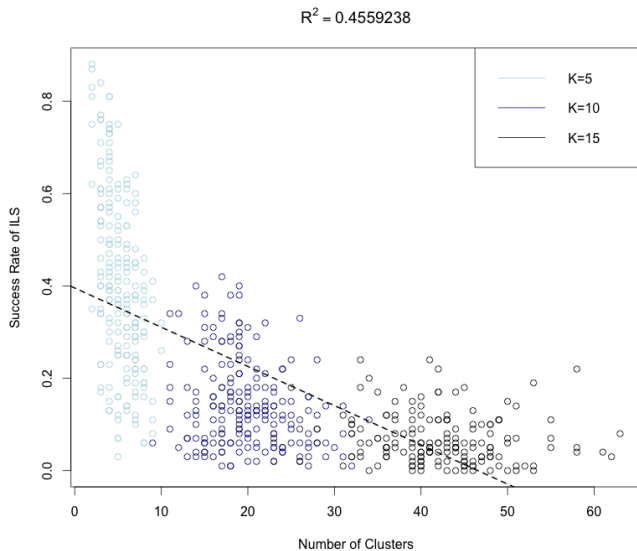$R^2 = 0.9430232$ (right figure)

Figure 3: ILS success rate over the number of clusters. Each dot represents a particular NK landscape. The dashed line is a univariate linear regression model, which is identical to the Pearson correlation.
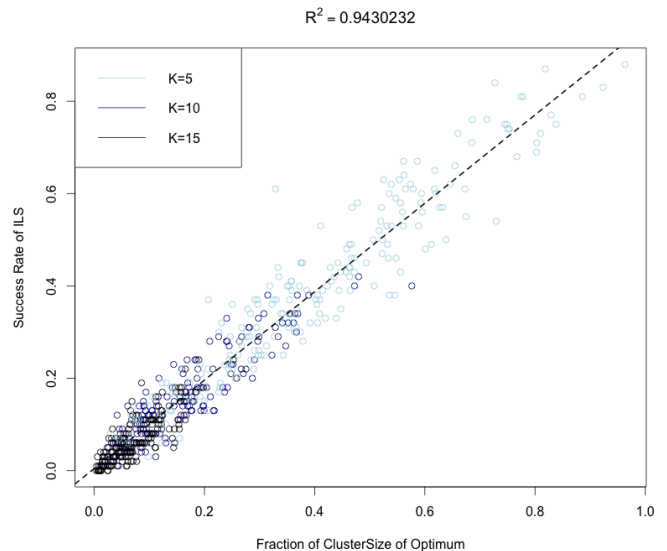
Figure 4: ILS success rate over relative size of the cluster which contains the global optimum. Each dot represents a particular NK landscape. The dashed line is a univariate linear regression model, which is identical to the Pearson correlation.

optimum.

Interestingly, we found that the center of the stars are the final local optima that are returned by ILS. Thus, the global optimum is always a center of a star and among the group of attractors. This finding complements earlier results by which we found that the PageRank centrality of a LON reflects the stationary distribution of both pure local search algorithms [12] and ILS [11]. Our results indicate that the relative cluster size of the attractors constructed by MCL is nearly identical to the PageRank value of the absorbing-state local optima in the landscape. This observation could offer an alternative way for performance prediction, however, we leave this for further research.

6. CONCLUSION

We conducted an experimental study on the characterization of a multi-funnel structure in fitness landscapes emerging from the Kauffman NK model under iterated local search. To analyze the presence of funnels in the landscapes, we used local optima networks and applied an algorithm for community detection, i.e. the Markov Cluster Algorithm.

The results confirm that the resulting landscapes consist of several clusters (communities), and the number of clusters grows with the number of interdependencies between the decision variables (epistasis). A higher number of clusters leads to a higher search difficulty, measured by the empirical success rate of ILS. An explanation for this observation is that ILS gets stuck due to the presence of funnels, which cannot be overcome by the perturbation operator applied. We estimate that a stronger perturbation operator could be used in such a case to overcome this situation.

Furthermore, the size of the cluster which contains the global optimum is strongly correlated to the success rate of ILS. This is no surprise when considering that—given a fixed number of local optima—a higher number of clusters

should in general lead to smaller clusters (by the number of nodes) and thus to higher search difficulty. Another explanation is that smaller clusters are generally harder to locate, since they require ILS to perform more perturbation steps. Thus, the probability to find the global optimum's cluster decreases with lower size of this cluster. In summary, the global optimum can be found more easily when it is surrounded by many other local optima.

The size of a cluster returned by the Markov cluster algorithm also offers a new possibility to predict the performance of ILS. We conjecture that the effect of MCL is closely related to the PageRank Centrality, which we leave for further research. Another finding that could be interesting for further research is that there are nodes connecting different areas in the fitness landscape. It would be interesting to see if this could be exploited by a search operator to help escape from one cluster to another.

In summary, our analysis has shown by community detection in LONs that the underlying landscapes are clustered, and that the presence and shape of these clusters are related to search difficulty. Furthermore, the results obtained by community detection have sufficient quality in terms of the modularity measure. We conclude that communities in LONs are a novel way to characterize the notion of funnels in fitness landscapes.

7. REFERENCES

[1] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, jan 2002.

[2] D. Applegate, W. Cook, and A. Rohe. Chained Lin-Kernighan for Large Traveling Salesman Problems. *INFORMS Journal on Computing*, 15(1):82–92, 2003.

[3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175–308, feb 2006.

[4] K. D. Boese, A. B. Kahng, and S. Muddu. A new adaptive multi-start technique for combinatorial global optimizations. *Operations Research Letters*, 16(2):101–113, 1994.

[5] S. P. Borgatti. Centrality and network flow. *Social Networks*, 27(1):55–71, jan 2005.

[6] F. Daolio, M. Tomassini, S. Vérel, and G. Ochoa. Communities of minima in local optima networks of combinatorial spaces. *Physica A: Statistical Mechanics and its Applications*, 390(9):1684–1694, may 2011.

[7] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, feb 2010.

[8] L. C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239, 1979.

[9] F. Glover. Future paths for integer programming and links to artificial intelligence. *Computers & Operations Research*, 13(5):533–549, jan 1986.

[10] D. R. Hains, L. D. Whitley, and a. E. Howe. Revisiting the big valley search space structure in the TSP. *Journal of the Operational Research Society*, 62(2):305–312, 2011.

[11] S. Herrmann. Determining the Difficulty of Landscapes by PageRank Centrality in Local Optima Networks. In *European Conference on Evolutionary Computation in Combinatorial Optimization (EvoCOP 2016)*, pages 74–87. Porto, 2016.

[12] S. Herrmann and F. Rothlauf. Predicting Heuristic Search Performance with PageRank Centrality in Local Optima Networks. In *Proceedings of the 2015 Genetic and Evolutionary Computation Conference - GECCO '15*, pages 401–408, Madrid, Spain, 2015. ACM Press.

[13] S. Kauffman and S. Levin. Towards a General Theory of Adaptive Walks on Rugged Landscapes. *Journal of theoretical Biology*, pages 11–45, 1987.

[14] P. Kerschke, M. Preuss, S. Wessing, and H. Trautmann. Detecting Funnel Structures by Means of Exploratory Landscape Analysis. In *Proceedings of the 2015 Genetic and Evolutionary Computation Conference - GECCO '15*, pages 265–272, New York, New York, USA, 2015. ACM Press.

[15] S. Lin and B. W. Kernighan. An Effective Heuristic Algorithm for the Traveling-Salesman Problem. *Operations Research*, 21(2):498–516, apr 1973.

[16] M. Locatelli. On the Multilevel Structure of Global optimization problems. *Computational Optimization and Applications*, 30(1):5–22, 2005.

[17] H. R. Lourenço, O. C. Martin, and T. Stützle. Iterated Local Search. In *Handbook of Metaheuristics*, pages 320–353. Kluwer Academic Publishers, Boston, 2003.

[18] M. Lunacek and D. Whitley. The dispersion metric and the CMA evolution strategy. In *Proceedings of the 8th annual conference on Genetic and evolutionary computation - GECCO '06*, page 477, New York, New York, USA, 2006. ACM Press.

[19] C. P. Massen and J. P. K. Doye. Identifying communities within energy landscapes. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 71(4):1–13, 2005.

[20] M. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of . . .* , 103(23):8577–82, 2006.

[21] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. pages 1–16, 2003.

[22] G. Ochoa, M. Tomassini, S. Vérel, and C. Darabos. A study of NK landscapes' basins and local optima networks. In *Proceedings of the 10th annual conference on Genetic and evolutionary computation - GECCO '08*, page 555, New York, 2008. ACM Press.

[23] G. Ochoa and N. Veerapen. Deconstructing the Big Valley Search Space Hypothesis. In *European Conference on Evolutionary Computation in Combinatorial Optimization (EvoCOP 2016)*, pages 58–73. LNCS, Springer, Porto, 2016.

[24] G. Ochoa, N. Veerapen, D. Whitley, and E. K. Burke. The Multi-Funnel Structure of TSP Fitness Landscapes: A Visual Exploration. In *Artificial Evolution (EA 2015)*, pages 1–13. Lyon, 2016.

[25] M. a. Porter, J.-P. Onnela, and P. J. Mucha. Communities in Networks. *Notices of the AMS*, 486(3-5):1082–1097, 2009.

[26] F. Rothlauf. *Design of modern heuristics: Principles and application.* Springer, Berlin and Heidelberg, 2011.

[27] V. Satuluri, S. Parthasarathy, and D. Ucar. Markov clustering of protein interaction networks with improved balance and scalability. In *Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology - BCB '10*, page 247, New York, USA, 2010. ACM Press.

[28] F. H. Stillinger. A Topographic View of Supercooled Liquids and Glass Formation. *Science*, 267(5206):1935–1939, mar 1995.

[29] E. Talbi and P. Bessière. A parallel genetic algorithm for the graph partitioning problem. In *Proceedings of the 5th international conference on Supercomputing - ICS '91*, pages 312–320, New York, USA, 1991. ACM Press.

[30] T. W. Valente. Network models of the diffusion of innovations. *Computational and Mathematical Organization Theory*, 2(2):134, 1996.

[31] S. van Dongen. *Graph clustering by flow simulation.* PhD thesis, Utrecht University, 2001.

[32] S. Vérel, F. Daolio, G. Ochoa, and M. Tomassini. Local optima networks with escape edges. In *Artificial Evolution*, pages 49–60, Angers, France, 2012. Springer.

[33] E. Weinberger. Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biological cybernetics*, 336:325–336, 1990.

[34] S. Wright. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In *Proceedings of the 6th International Congress of Genetics*, pages 356–366, 1932.