The Component Model for Elementary Landscapes and Partial Neighborhoods

Darrell Whitley

Computer Science, Colorado State University, Fort Collins, CO USA

Andrew M. Sutton

Computer Science, Colorado State University, Fort Collins, CO USA

Gabriela Ochoa

Computing Science and Mathematics, University of Stirling, Stirling, Scotland

Francisco Chicano

Lenguajes y Ciencias de la Computación, University of Málaga, Málaga, Spain

Abstract

Local search algorithms exploit moves on an adjacency graph of the search space. An "elementary landscape" exists if the objective function f is an eigenfunction of the Laplacian of the graph induced by the neighborhood operator; this allows various statistics about the neighborhood to be computed in closed form. A new component based model makes it relatively simple to prove that certain types of landscapes are elementary. The traveling salesperson problem, weighted graph (vertex) coloring and the minimum graph bisection problem yield elementary landscapes under commonly used local search operators. The component model is then used to efficiently compute the mean objective function value over partial neighborhoods for these same problems. For a traveling salesperson problem over n cities, the 2-opt neighborhood can be decomposed into $\lfloor n/2 - 1 \rfloor$ partial neighborhoods. For graph coloring and the minimum graph bisection problem, partial neighborhoods

Email addresses: `whitley@cs.colostate.edu` (Darrell Whitley),
`sutton@cs.colostate.edu` (Andrew M. Sutton), `gabriela.ochoa@cs.stir.ac.uk`
(Gabriela Ochoa), `chicano@lcc.uma.es` (Francisco Chicano)

can be used to focus search on those moves that are capable of producing a solution with a strictly improving objective function value.

---

1. Introduction

A *fitness landscape* for a combinatorial problem instance is defined by a triple (X, N, f). In this definition, X is a set of *candidate solutions*, and f is an *objective function* $f : X \to \mathbb{R}$ that maps each candidate solution to a real value. The objective is to either minimize or maximize f. Fitness landscapes are typically associated with local search methods that use a neighborhood move operator to define adjacency between points in the search space. We define a *neighborhood operator* as a function N that maps candidate solutions in X to subsets of X. Given a candidate solution $x \in X$, $N(x)$ is the set of points in X (i.e., the neighbors of x) that are adjacent to x. We say that a neighborhood is regular if the number of neighbors of each solution $x \in X$ is same and we denote with $d = |N(x)|$ this number. We say that a neighborhood is symmetric if the neighborhood relationship is symmetric, that is, $x \in N(y)$ if and only if $y \in N(x)$.

Grover [6] originally showed that for certain NP-hard problems there exist landscapes where it is possible to compute the mean objective function value over the set of neighbors $N(x)$ without explicitly evaluating any of the neighbors of x. He showed there exist neighborhoods for the traveling salesperson problem, graph coloring, minimum graph bisection, weight partitioning, as well as not-all-equal satisfiability where this calculation is possible. Stadler [13] named this class of problems "elementary landscapes" and Stadler showed that for these problems the objective function f is an eigenfunction of the Laplacian of the graph induced by the neighborhood operator. It can also be shown that a landscape with a symmetric neighborhood operator is elementary if and only if the time series generated by a random walk on the landscape is an AR(1) process [14, 5].

Other problems, such as maximum k-satisfiability, NK-landscapes, subset sum, and the quadratic assignment problem [16, 2] can be shown to be expressible as a superposition of a small number of elementary landscapes. Maximum 3-satisfiability, taken with the traditional Hamming neighborhood, can be expressed as a superposition of three elementary landscapes; it is not

2

only possible to compute the mean, but also to compute other statistical moments in polynomial time, including variance, skew and kurtosis. These statistics can be computed in polynomial time even over generalized neighborhoods that are exponentially large [16].

Some elementary landscapes correspond to problems whose objective functions are linear combinations of components drawn from some finite set C. In these cases, the objective function can be characterized as a *discrete linear subset problem* introduced by Papadimitriou and Steiglitz [9]. Each candidate solution to such a problem is defined by some subset of $x \subseteq C$, and the objective function is a weighted sum over the components in x. A landscape is *elementary* when the set of candidate solutions and the objective function are coupled with a neighborhood operator that moves components in and out of solutions with uniform frequency. We give a rigorous definition of this *component model* in Section 2.1.

In this paper, we also answer the question of whether or not, using the component model, a neighborhood can be partitioned in such a way to explicitly calculate how components are sampled in the partitions of the neighborhood. We show how the component model can be used to derive conditions under which there will exist partial neighborhoods that retain some of the properties that characterize the full neighborhood. We give a formal definition of partial neighborhood in Section 2.1.

From a theoretical point of view, the existence of elementary landscapes and the ability to compute statistical information about neighborhoods and partial neighborhoods of elementary landscapes is inherently interesting. It is too soon to show that this information can be leveraged to build new and improved local search algorithms. But there have already been some breakthroughs. We can now find the improving moves in a unit distance Hamming neighborhood in $O(1)$ time without explicitly generating any of the neighbors; this result holds for NK-landscapes, maximum k-satisfiability, and all pseudo-Boolean functions where the objective function taken together with the Hamming neighborhood operator is a superposition of elementary landscapes [17]. We can also use the average of the neighborhood two moves ahead as a surrogate for the objective function. Thus, instead of optimizing the objective function f directly, we can instead optimize $\text{avg } N(x)$, which corresponds to the expected value of the next move, one move ahead. In some cases, this can yield better results than optimizing f directly [18].

From a practical point of view, the results we present on partial neighborhoods are useful because they provide statistical information about neigh-

3

borhood operators that make limited changes to a current solution. In some cases, it may be desirable to limit the changes of a local search move operator so that it is not too disruptive. One example is in the planar traveling salesperson problem (TSP) where a partial neighborhood corresponds to the set of all tour inversions of length ℓ (see Section 4.1). An instance of planar TSP is given by a set of points P in the Euclidean plane, and a candidate solution is a Hamiltonian circuit through P. Let $H \subseteq P$ denote the set of points that lie on the convex hull of P. If the points in H appear in a candidate solution in the same order they do on the convex hull, then their relative order is already correct. Indeed, it is possible to design randomized search heuristics that exploit this property [8]. Thus any inversion move that destroys this order is likely to be too disruptive for a local search operator. From such a solution, one may be interested in performing only short inversions of up to length $\ell = |P| - |H|$. This corresponds to a set of partial 2-opt neighborhoods that we introduce in this paper.

We are only just beginning to understand how new and more detailed information about the local search landscape graph can be exploited by search algorithms. The current paper provides not only a foundation for better understanding elementary landscapes, but also for extracting statistical information about partial neighborhoods.

In the next section we briefly review basic mathematical properties of elementary landscapes. We then formally introduce the "component model" for elementary landscapes, providing a more rigorous and solid formulation than the one presented in the previous work [19, 21]. Moreover, we approach the characterization from a new perspective that links the component model with the concept of discrete linear subset problems. In Section 3 we provide proofs that the traveling salesperson problem, the weighted graph (vertex) coloring, and the minimum graph bisection are elementary landscapes. These results are known and are included to make the paper self-contained. However, the proof in Section 3.2 for the weighted graph coloring is novel and more rigorous than the previous proof [21]. In Section 4 we analyze partial neighborhoods of the three problems. For each case we extend and generalize the previous results and provide alternative approaches that simplify the mathematical developments. We conclude the paper in Section 5.

2. Elementary Landscapes

For a fitness landscape (X, N, f), the neighborhood operator can be represented by a $|X| \times |X|$ *adjacency matrix*

$$\mathbf{A}_{xy} = \begin{cases} 1 & \text{if } y \in N(x); \\ 0 & \text{otherwise.} \end{cases}$$

We will restrict our attention to *regular neighborhoods*, where $|N(x)| = d$ for all $x \in X$. When a neighborhood is regular, the Laplacian operator can be defined as

$$\Delta = \mathbf{A} - d\mathbf{I},$$

where the Laplacian acts on the fitness function f as follows

$$\Delta f = \begin{bmatrix} \sum_{y \in N(x_1)} (f(y) - f(x_1)) \\ \sum_{y \in N(x_2)} (f(y) - f(x_2)) \\ \vdots \\ \sum_{y \in N(x_{|X|})} (f(y) - f(x_{|X|})) \end{bmatrix}.$$

The element of this matrix-vector product corresponding to a point x can thus be written as

$$\Delta f(x) = \sum_{y \in N(x)} (f(y) - f(x)). \tag{1}$$

A fitness landscape (X, N, f) is an *elementary landscape* if and only if f is an eigenfunction of the Laplacian of the adjacency matrix for N with a constant offset of b [14]. Usually b is proportional to $\bar{f}$, the mean fitness value in X. In particular, Grover's wave equation can be written as

$$\Delta f + k(f - \bar{f}) = 0,$$

where k is a positive constant. If we assume that f is normalized such that $\bar{f} = 0$ then it follows that

$$\Delta f + k(f - \bar{f}) = \Delta f + kf = 0 \quad \text{and therefore} \quad \Delta f = -kf.$$

When f does not have zero mean, we can use the equation

$$\Delta f(x) = k\bar{f} - kf(x),$$

to calculate the *average fitness* across the neighborhood of any given candidate solution x. Using Equation (1), we calculate this average fitness as follows:

$$
\begin{aligned}
\operatorname{avg} N(x) &= \frac{1}{d} \sum_{y \in N(x)} f(y) \\
&= \frac{1}{d} \sum_{y \in N(x)} (f(y) - f(x)) + f(x) \\
&= \frac{1}{d} \Delta f(x) + f(x) \\
&= f(x) + \frac{k}{d}(\bar{f} - f(x)).
\end{aligned}
\tag{2}
$$

2.1. Components and Partial Neighborhoods

In this section we introduce a formal treatment of the *component model* of elementary landscapes that will be used throughout the paper.

Let $\mathbb{R}_{\geq 0} = \{z \in \mathbb{R} \mid z \geq 0\}$. Following Rowe and Vose [11], we consider the following type of problem.

Definition 1. *A discrete linear subset problem* class is given by a set C of components, where each candidate solution, $x \in X$ is also a subset of C. The components in C have an associated weight function $w : C \rightarrow \mathbb{R}_{\geq 0}$, allowing the objective function to be expressed as a linear combination of the components in x:*

$$
f(x) = \sum_{i \in x} w(i).
$$

where i indexes the components in candidate solution x.

Many classical combinatorial optimization problems are discrete linear subset problems. In this paper, the components will be edges in a graph. However, components can be other structures such as k-CNF clauses, pairs of objects, etc.

We can regard a discrete linear subset problem algebraically by associating each element $x \in X$ along with the weight function as elements of $\mathbb{R}^{|C|}$. In particular, each set $x \in X$ can be represented by a characteristic 0-1 vector $\mathbf{x} \in \mathbb{R}^{|C|}$ where the i-th element of $\mathbf{x}$ is 1 if and only if $i \in x$, otherwise it is 0. Similarly, we can treat the weight function as a vector $\mathbf{w} \in \mathbb{R}^{|C|}$. The objective function for any discrete linear subset problem is defined in terms of an inner product between a characteristic vector and the weight vector,

i.e., $f(x) = \mathbf{x}^\top \mathbf{w}$. Thus, any discrete linear subset problem can be uniquely defined in terms of $(X, \mathbf{w})$.

We will use the following notational convention for a logical expression P.

$$[P] = \begin{cases} 1 & \text{if the expression } P \text{ is true;} \\ 0 & \text{otherwise.} \end{cases}$$

We also adopt the set theoretic convention of denoting the power set of a set S as 2^S.

We can now characterize elementary landscapes completely in terms of the discrete linear subset problem. Formally, this is the *component model* of an elementary landscape.

Theorem 1 (Component Theorem). *Suppose $(X, \mathbf{w})$ is a discrete linear subset problem where the sum of the weights $\sum_{i \in C} w(i)$ is different than zero. We say a symmetric and regular neighborhood operator $N : X \to 2^X$ induces an elementary landscape (X, N, f) if, for any component $i \in C$ and all $x \in X$,*

$$\sum_{y \in N(x)} [i \in x][i \notin y] = \alpha[i \in x], \quad \text{and} \quad \sum_{y \in N(x)} [i \notin x][i \in y] = \beta[i \notin x], \quad (3)$$

where α and β are fixed scalars. Note that the ratios $p_1 = \alpha/d$, $p_2 = \beta/d$ and $p_3 = \bar{f} \cdot \left(\sum_{i \in C} w(i) \right)^{-1}$ characterize the elementary landscape.

Proof. For an arbitrary $x \in X$ and its corresponding characteristic vector $\mathbf{x} \in \mathbb{R}^{|C|}$ we have

$$\mathbf{x}_i = [i \in x] \quad \text{and} \quad 1 - \mathbf{x}_i = [i \notin x].$$

Thus, Equation (3) can be written as

$$\sum_{y \in N(x)} \mathbf{x} \odot (\mathbf{1} - \mathbf{y}) = \alpha \mathbf{x}, \quad \text{and} \quad \sum_{y \in N(x)} (\mathbf{1} - \mathbf{x}) \odot \mathbf{y} = \beta (\mathbf{1} - \mathbf{x}), \quad (4)$$

where $\mathbf{1} = (1, 1, \ldots, 1)^\top$ and $\odot$ denotes elementwise multiplication (Hadamard product).

We have the following identity:

$$\mathbf{y} = \mathbf{y} + (\mathbf{x} - \mathbf{x}) + (\mathbf{x} \odot \mathbf{y} - \mathbf{x} \odot \mathbf{y}) = \mathbf{x} - \mathbf{x} \odot (\mathbf{1} - \mathbf{y}) + (\mathbf{1} - \mathbf{x}) \odot \mathbf{y}. \quad (5)$$

7

Hence,

$$\sum_{y \in N(x)} f(y) = \sum_{y \in N(x)} \mathbf{y}^\top \mathbf{w}$$

$$= \sum_{y \in N(x)} (\mathbf{x} - \mathbf{x} \odot (\mathbf{1} - \mathbf{y}) + (\mathbf{1} - \mathbf{x}) \odot \mathbf{y})^\top \mathbf{w} \quad \text{by (5)}$$

$$= \sum_{y \in N(x)} \mathbf{x}^\top \mathbf{w} - \left(\sum_{y \in N(x)} \mathbf{x} \odot (\mathbf{1} - \mathbf{y}) \right)^\top \mathbf{w} \tag{6}$$

$$+ \left(\sum_{y \in N(x)} (\mathbf{1} - \mathbf{x}) \odot \mathbf{y} \right)^\top \mathbf{w}$$

$$= |N(x)| \mathbf{x}^\top \mathbf{w} - (\alpha \mathbf{x})^\top \mathbf{w} + (\beta(\mathbf{1} - \mathbf{x}))^\top \mathbf{w} \quad \text{by (4)}$$

$$= d \cdot f(x) - \alpha f(x) + \beta(\mathbf{1}^\top \mathbf{w} - f(x)).$$

It is now straightforward to state the average objective function value over the neighborhood in terms of the components using the ratios p_1, p_2, and p_3 defined above. The average objective function value over the neighborhood is thus

$$\text{avg } N(x) = \frac{1}{d} \sum_{y \in N(x)} f(y)$$

$$= f(x) - p_1 f(x) + p_2(\mathbf{1}^\top \mathbf{w} - f(x)) \quad \text{by (6)} \tag{7}$$

$$= f(x) - (p_1 + p_2) f(x) + (p_2/p_3) \bar{f}$$

$$= f(x) + \frac{\alpha + \beta}{d}(\bar{f} - f(x)),$$

where we have used the (still unproven) expression $p_1 + p_2 = p_2/p_3 = (\alpha + \beta)/d$. Thus we have recovered the difference equation in Equation (2) with $k = \alpha + \beta$. All elementary landscapes in the component model are parameterized by the three ratios p_1, p_2, and p_3. These ratios are constant across every $N(x)$ with $x \in X$, and p_3 is always independent of the neighborhood operator and the neighborhood size.

Now we have to prove that $p_1 + p_2 = p_2/p_3$, or equivalently, $p_3 = p_2/(p_1 + p_2) = \beta/(\alpha + \beta)$. In order to do this we introduce a component matrix $\mathbf{C}$ of dimension $|X| \times |C|$ and indexed by the solutions such that row x is the characteristic vector $\mathbf{x}$. If we multiply this matrix by the Laplacian Δ we

8

have:

$$\Delta\mathbf{C} = -\alpha\mathbf{C} + \beta(\mathbb{1} - \mathbf{C}), \tag{8}$$

where $\mathbb{1}$ denotes a $|X| \times |C|$ matrix with all ones. Equation (8) holds since each row x of $\Delta\mathbf{C}$ is counting with positive numbers the components not in x that are included in the neighboring solutions of x and with negative numbers the components in x that are removed from the solution to generate the neighbors. Now we multiply (8) by the row vector $\mathbf{1}^\top$ from the left to obtain:

$$\mathbf{1}^\top\Delta\mathbf{C} = -\alpha\mathbf{1}^\top\mathbf{C} + \beta\mathbf{1}^\top(\mathbb{1} - \mathbf{C}).$$

The product $\mathbf{1}^\top\Delta$ is zero, since the neighborhood is symmetric. In effect, from the definition of Δ we have $\Delta\mathbf{1} = 0$, and for symmetric neighborhoods Δ is symmetric, so $(\mathbf{1}^\top\Delta)^\top = \Delta^\top\mathbf{1} = \Delta\mathbf{1} = 0$. On the other hand, the row vector $\mathbf{1}^\top\mathbf{C}$ gives the number of times that each component appears in solutions over the entire search space. By abusing notation slightly, we write the column vector $\mathbf{p_3} = (\mathbf{1}^\top\mathbf{C})^\top/|X|$. With all these considerations we can write:

$$-\alpha\mathbf{1}^\top\mathbf{C} + \beta\mathbf{1}^\top(\mathbb{1} - \mathbf{C}) = -\alpha|X|\mathbf{p_3}^\top + \beta|X|\mathbf{1}^\top - \beta|X|\mathbf{p_3}^\top = 0,$$

and solving for $\mathbf{p_3}$ we have:

$$\mathbf{p_3} = \frac{\beta}{\alpha + \beta}\mathbf{1}.$$

This means that all the components appear the same number of times in the entire search space and the parameter p_3 is given by $\beta/(\alpha + \beta)$. This concludes the proof. $\square$

Observing Equation (3) more closely, an elementary landscape can be characterized by a discrete linear subset problem along with a neighborhood operator that, for all $x \in X$,

1. Every component $i \in x$ is removed exactly α times across the neighborhood $N(x)$.
2. Every component $i \notin x$ is appears in a neighbor $y \in N(x)$ exactly β times across the neighborhood $N(x)$.

There are some landscapes in which the solution x contains always the same number of components, $|x|$, and the neighborhood operator replaces

9

always the same number of components, ρ, from the solution. Two examples are the TSP with the 2-opt neighborhood (Section 3.1) and the minimum graph bisection problem with the swap neighborhood (Section 3.3). When this happens the ratios p_1, p_2 and p_3 can be written in an alternative way. The number of times that a component is removed from the solution, α, multiplied by the number of components in one solution, $|x|$, must be equal to the size of the neighborhood, d, multiplied by the number of components removed from the solution for each neighbor, ρ. That is: $|x|\alpha = d\rho$. Then p_1 can be written as:

$$p_1 = \frac{\alpha}{d} = \frac{\rho}{|x|}.$$

A similar argument can be used for the component inserted in the solution x to form a neighbor and we have: $|C \setminus x|\beta = d\rho$. Then we can write:

$$p_2 = \frac{\beta}{d} = \frac{\rho}{|C \setminus x|}.$$

The third ratio, p_3, can also be written in a different form:

$$p_3 = \frac{\bar{f}}{\sum_{i \in C} w(i)} = \frac{\sum_{x \in X} \sum_{i \in x} w(i)}{|X| \sum_{i \in C} w(i)} = \frac{\sum_{i \in C} w(i)|\{x \in X | i \in x\}|}{|X| \sum_{i \in C} w(i)}.$$

We now take into account that in elementary landscapes with a symmetric and regular neighborhood any component, $i \in C$, must appear the same number of times, η, in the complete search space, that is, the term $|\{x \in X | i \in x\}| = \eta$ is independent of i. Furthermore, since each solution has $|x|$ components we have $|x||X| = \eta|C|$ and we can write:

$$p_3 = \frac{\bar{f}}{\sum_{i \in C} w(i)} = \frac{\eta}{|X|} = \frac{|x|}{|C|}.$$

It is interesting to look at cases where different neighborhood operators induce elementary landscapes for the same problem class. A motivating example is the symmetric traveling salesperson problem (TSP). Indeed, the TSP is elementary under both the 2-exchange operator [6] and the 2-opt operator [21]. These operators exchange components differently, and result in different sizes of neighborhood.

Suppose we have two different neighborhood operators N and N' that both induce an elementary landscape from some discrete linear subset problem class. In particular, suppose α and β are the constants in Equation (6)

for N, and α' and β' are the constants corresponding to N'. Since both N and N' operators induce an elementary landscape for the same problem class, then

$$\bar{f} = p_3 \sum_{i \in C} w(i) = \frac{p_2}{p_1 + p_2} \sum_{i \in C} w(i) = \frac{\beta}{\alpha + \beta} \sum_{i \in C} w(i) = \frac{\beta'}{\alpha' + \beta'} \sum_{i \in C} w(i).$$

Since $\beta/(\alpha + \beta) = \beta'/(\alpha' + \beta')$ it follows that the kinds of neighborhood operators that induce elementary landscapes are restricted, and the component model illuminates the reasons why.

We may now also formally define the partial neighborhood concept as follows.

Definition 2. *Let $N : X \to 2^X$ be a neighborhood on X. A* partial neighborhood *with respect to N is a neighborhood operator $N' : X \to 2^X$ such that, for all $x \in X$, $N'(x) \subseteq N(x)$.*

In many cases, we can define a partial neighborhood with properties that are similar to the ones in the claim of Theorem 1. In this paper, we will investigate partial neighborhoods N' with respect to neighborhoods N where N induces an elementary landscape (X, N, f). For all partial neighborhoods we investigate, there is a fixed scalar α' such that, for all $x \in X$, and for all $i \in C$,

$$\sum_{y \in N'(x)} [i \in x][i \notin y] = \alpha'[i \in x].$$

This will allow for the calculation of the average objective function value over the neighborhood. In particular, in each case we will also derive a β' and a g such that

$$\operatorname{avg} N'(x) = f(x) - \frac{\alpha'}{d'} f(x) + \frac{\beta'}{d'} g(C \setminus x),$$

where g is a term (locally depending on x) that is determined by how the components in a subset of $C \setminus x$ appear in the neighborhood x.

With this calculation, it is also possible to add together the contributions of the partial neighbors into combinations of partial neighborhoods, and to find the averages over the complements of partial neighborhoods. Let N be any neighborhood that induces an elementary landscape on a discrete linear subset problem $(X, \mathbf{w})$. Let N' and N'' be two partial neighborhoods with respect to N such that, for all $x \in X$,

1. $|N'(x)| = d'$ and $|N''(x)| = d''$,
2. $N'(x) \cap N''(x) = \emptyset$.

For all $x \in X$, define N^* as $N^*(x) = N'(x) \cup N''(x)$, and define $\overline{N'}$ as $N(x) \setminus N'(x)$. We have

$$\text{avg } N^*(x) = (d'(\text{avg } N'(x)) + d''(\text{avg } N''(x)))/d^*,$$

where $d^* = d' + d''$ and

$$\text{avg } \overline{N'}(x) = (d(\text{avg } N(x)) - d'(\text{avg } N'(x)))/(d - d').$$

3. Demonstrating that Landscapes are Elementary

We first use the component model to show that the traveling salesperson problem, graph coloring, and the minimum graph bisection problem are elementary under a commonly used operator.

3.1. The Traveling Salesperson Problem

The classic 2-opt operator for the traveling salesperson problem (TSP) cuts two edges in a Hamiltonian circuit corresponding to a tour; this breaks the Hamiltonian circuit into two segments. One of these segments is reversed, and the segments are reattached. Suppose N corresponds to the 2-opt neighborhood operator. If x is the current solution and $y \in N(x)$, then 2-opt breaks 2 edges in the tour x, and adds two new edges to create y.

Stadler and Schnabl [15] showed that an "inversion operator" which is a superset of the 2-opt operator induces an elementary landscape. The proof presented here more precisely describes the classic 2-opt neighborhood used by the Lin and Kernighan (1973) local search algorithm for the TSP [7].

Let $G = (V, E)$ be an undirected graph on $|V| = n$ vertices and let $w : E \to \mathbb{R}_{\geq 0}$ be a weight function on the edges. Thus, the TSP is a discrete linear subset problem as defined in Definition 1 where the component set $C = E$ is the set of $n(n-1)/2$ edges and w specifies the problem instance by giving the cost of each edge.

We first establish that all edges occur with equal frequency across all possible solutions. Every component $i \in C$ is an edge $i = (u, v)$ that can be used as the *initial edge* in a circuit starting at city u followed immediately by city v. For every initial edge, there are $n - 2$ unvisited cities which can be used to construct a circuit, so every initial edge must occur in the same

number of tours. Since there are n edges in a solution and edges appear with uniform frequency, it follows that over all possible tours:

$$p_3 = \frac{n}{|C|} = \frac{n}{n(n-1)/2} = \frac{2}{n-1},$$

$$\bar{f} = p_3 \sum_{i \in C} w(i) = \frac{2}{n-1} \sum_{i \in C} w(i).$$

A solution x is a set of edges that correspond to a Hamiltonian tour in the graph G. To determine p_1, note there are n edges in any solution x, and a single 2-opt move changes exactly two edges. Since removing adjacent edges does not result in a valid 2-opt move, it follows that, over the entire neighborhood, each edge in x is removed exactly $n - 3$ times. Therefore

$$p_1 = 2/n \quad \text{and} \quad p_1 = \frac{2(n-3)/2}{n(n-3)/2} = \frac{\alpha}{d}.$$

There are $|C| - n$ edges in $C \setminus x$. The edges that reconnect the Hamiltonian tour to form a valid 2-opt neighbor of x must come from this set. Again, the edges in $C \setminus x$ are uniformly sampled by cutting at every feasible location; all pairs of cities that are not adjacent in x are adjacent in some $y \in N(x)$. Thus,

$$p_2 = \frac{2}{(n(n-1)/2) - n} = \frac{2}{n(n-3)/2} = \frac{\beta}{d}.$$

These terms can simply be substituted into Equation (7) to yield the average objective function value over the neighborhood.

$$\text{avg } N(x) = f(x) + \frac{\alpha + \beta}{d}(\bar{f} - f(x)) = f(x) + \frac{n-1}{n(n-3)/2}(\bar{f} - f(x)).$$

A good deal of the work on elementary landscapes has focused on the TSP. This includes a proof that the symmetric TSP is elementary under 2-exchange [6] and 2-opt and 3-exchange [3] The weakly-symmetric TSP [12], and variants of the multiple TSP are also elementary [4].

3.2. Weighted Graph Coloring

Grover showed that graph coloring (vertex coloring) yields an elementary landscape under an operator that recolors every vertex; if the number of colors is r, recoloring a vertex uses the other $r - 1$ colors. Given r colors,

the goal is to assign a color to every vertex of graph G so as to minimize the total cost of the coloring. Grover assumes a conflicting coloring is assigned unit cost. We generalize Grover's result to show that weighted graph coloring problem is elementary under the same operator [2]. Many practical combinatorial optimization problems reduce to weighted graph coloring. Examples include the basic Frequency Assignment problem [20], the basic Time Tabling problem [10], and register allocation [1].

Let $G = (V, E)$ be the complete graph on $|V| = n$ vertices and let us define $w : E \to \mathbb{R}_{\geq 0}$ to be a nonnegative weight function on the edges.

A coloring is a map from V to a color set $\{1, 2, \ldots, r\}$. The objective is to find a coloring that minimizes the cost of edges that have both endpoints assigned to the same color. An edge that has both endpoints assigned to the same color in some coloring is called a *conflict* for that coloring. The set of edges that are a conflict for a coloring is called the *conflict set*.

Let $Z = \{1, 2, \ldots, r\}^n$ be the set of all r-colorings on V. Again, we let the component set C be equal to the edge set E of G. The set of candidate solutions is the collection X of conflict sets for all colorings. In the case of weighted graph coloring, it is necessary to be more careful because we lose information moving from Z into X. In particular, the map from Z to X is not injective. For example, it is possible for two distinct colorings $z_1, z_2 \in Z$ to have distinct neighborhoods, (e.g., a neighbor of z_1 is not a neighbor of z_2 or vice versa) but identical conflict sets. A candidate solution in the landscape therefore cannot be uniquely identified by its conflict set. Figure 1 illustrates this problem.

To resolve this, we must augment the component model slightly. The proof of the following theorem follows directly from Theorem 1.

Theorem 2. *Let Z and C be finite sets. Define $\chi : Z \to 2^C$ to be a map from Z into collections of subsets of C. Let X be the image of Z under χ and assume that $(X, \mathbf{w})$ is a discrete linear subset problem as defined in Definition 1. Then a neighborhood operator $N : Z \to 2^Z$ induces an elementary landscape $(Z, N, (f \circ \chi))$ if, for any component $i \in C$ and all $z \in Z$,*

$$\sum_{y \in N(z)} [i \in \chi(z)][i \notin \chi(y)] = \alpha[i \in \chi(z)], \quad and,$$

$$\sum_{y \in N(z)} [i \notin \chi(z)][i \in \chi(y)] = \beta[i \notin \chi(z)], \tag{9}$$

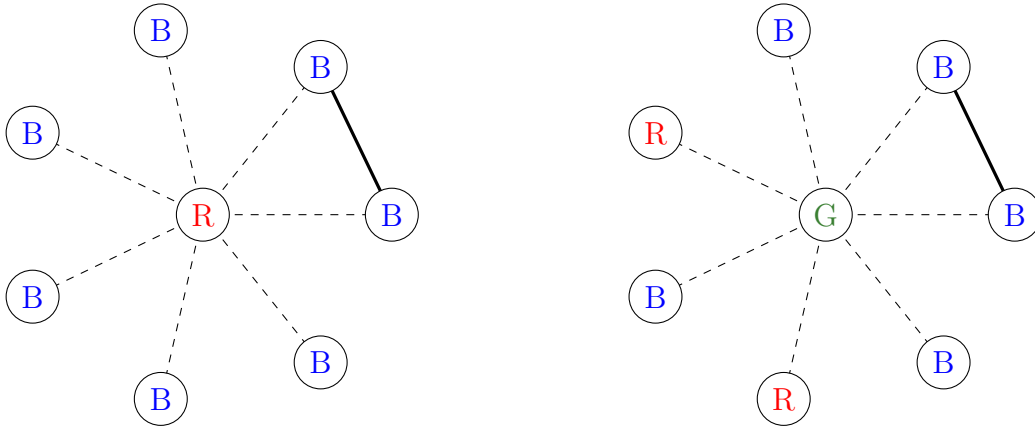*where α and β are fixed scalars.*

14

Figure 1: Two colorings with the same conflict set (invisible edges have zero weight, solid edge denotes conflict). Recoloring the center vertex in the graph on the right creates a neighbor that doesn't exist in the recoloring neighborhood of the graph on the left.

Obviously, the difference equations introduced in Section 2.1, i.e., (6) and (7), also hold for this augmented component model. Furthermore, one may consider the original component model with the above definition by letting χ be the identity map.

In the case of vertex coloring, let $\chi : Z \to X$ be defined as

$$\chi(z) = \{(u, v) \in C \mid z(u) = z(v)\},$$

that is, the conflict set of z. The objective function is the weighted cost of a coloring, that is

$$f(\chi(z)) = \sum_{i \in \chi(z)} w(i).$$

Hence, the image of Z under χ taken together with the weight function on C is a discrete linear subset problem.

The neighborhood operator can now be defined as $N : Z \to 2^Z$ where

$$N(z) = \{y \in Z \mid d_H(z, y) = 1\}.$$

Here d_H is the Hamming distance between two colorings. Intuitively, $N(z)$ is the set of all colorings that differ by one vertex from z; this corresponds to the recoloring neighborhood mentioned above. Any two colorings are neighbors if they can be transformed into one another by recoloring exactly one vertex. Clearly, $|N(z)| = d = |V|(r - 1)$.

We now show that the augmented component model given by weighted graph coloring taken together with the recoloring neighborhood N is an elementary landscape. Let $(u, v) \in C$ be an arbitrary component. There are r^2 ways to color u and v, but only r such colorings have (u, v) in their conflict set. Thus each edge (u, v) is in the conflict set of a $r/r^2 = 1/r$ fraction of all colorings. This implies $p_3 = 1/r$ and the average objective function value over all solutions is

$$\bar{f} = p_3 \sum_{i \in C} w(i) = 1/r \sum_{i \in C} w(i).$$

We now consider a coloring z. Let $x = \chi(z)$ be the corresponding conflict set for z. Suppose $(u, v) \in x$. In this case, $z(u) = z(v)$ and there are $r - 1$ ways to recolor u or v that removes (u, v) from the conflict set. It follows that there are $2(r - 1)$ neighbors $y \in N(z)$ such that $\chi(y)$ does not contain the component (u, v) and we can define

$$p_1 = \frac{2(r - 1)}{|V|(r - 1)} = \frac{\alpha}{d}.$$

Similarly, consider $(u, v) \notin x$. In this case $z(u) \neq z(v)$, and a conflict is generated when u is colored the same as v or v is colored the same as u. Since there are only two ways this can happen, it follows that there are exactly two neighbors $y \in N(z)$ such that $\chi(y)$ does contain (u, v) and we can define

$$p_2 = \frac{2}{|V|(r - 1)} = \frac{\beta}{d}.$$

Substituting these terms into Equation (7) yields the average objective function value over the recoloring neighborhood for weighted vertex coloring.

$$\text{avg } N(z) = f(\chi(z)) - p_1 f(\chi(z)) + p_2(\bar{f}/p_3 - f(\chi(z)))$$
$$= f(\chi(z)) + \frac{2r}{|V|(r - 1)}(\bar{f} - f(\chi(z))).$$

This form satisfies Grover's wave equation with constant $k = 2r$ and neighborhood size $d = |V|(r - 1)$.

3.3. Minimum Graph Bisection

Let $G = (V, E)$ be the complete undirected graph on $|V| = n$ vertices such that n is even. Let $w : E \to \mathbb{R}_{\geq 0}$ be a weight function on the edges

of G. The minimum graph bisection problem is to find a bipartition of the vertices in G into a left hand side (LHS) and a right hand side (RHS) such that $|\text{LHS}| = |\text{RHS}| = n/2$ and the weighted sum of the edges connecting vertices in the LHS and RHS partitions is minimized.

A candidate solution is a set x of edges (u, v) where, without loss of generality, $u \in \text{LHS}$ and $v \in \text{RHS}$, such that LHS and RHS form a valid bipartition of G. A natural neighborhood operator is the *swap* operator that exchanges a vertex in the LHS with a vertex in the RHS. Since there are $n/2$ vertices in the LHS and $n/2$ in the RHS, for an arbitrary candidate solution x, there are $(n/2)^2 = n^2/4$ neighbors of x under the swap operator.

The set C of components is simply the set of $n(n - 1)/2$ edges, and this, together with the weight function, yields a discrete linear subset (see Definition 1). We now show that the swap operator induces an elementary landscape.

We begin by computing p_3: the average contribution to the objective function from an arbitrary component. Let $i = (u, v)$ be an arbitrary edge. If u is in one partition, then the fraction of total configurations where v is in the opposite partition is $\frac{n/2}{n-1}$. Assume the RHS has $n/2$ bins where a vertex can be placed, and that the LHS has $n/2$ bins where a vertex can be placed. After u is placed, there are $n - 1$ bins left in which v can be placed, where $n/2$ of these bins will be in the partition opposite of where u was placed.

$$p_3 = \frac{n/2}{n - 1} = \frac{n^2/4}{|C|},$$

$$\bar{f} = p_3 \sum_{i \in C} w(i) = \frac{n}{2(n - 1)} \sum_{i \in C} w(i).$$

Let x be an arbitrary candidate solution. To compute p_1, consider a component $i \in x$. By definition $i = (u, v)$ where (without loss of generality), $u \in \text{LHS}$ and $v \in \text{RHS}$. A swap operation can produce a neighbor that does not contain i by moving u to the RHS and keeping v in the RHS (by moving any of the other $(n/2 - 1)$ vertices in the RHS to the LHS). By symmetry, a swap operation can also produce a neighbor not containing i by moving v to the and keeping u in the LHS in $(n/2 - 1)$ ways. In total, there are $2(n/2 - 1)$ swap operations that produce a neighbor $y \in N(x)$ that does not contain component $i \in x$. Since there are $n^2/4$ total neighbors:

$$p_1 = \frac{2(n/2 - 1)}{n^2/4} = \frac{n - 2}{n^2/4} = \frac{\alpha}{d}.$$

Now, we consider a component $i \notin x$. By definition $i = (a, b)$ where a and b are in the same partition. A swap operation can produce a neighbor containing i by swapping a with a vertex in the opposite partition, or by swapping b with a vertex in the opposite partition. There are $n/2$ vertices on the opposite side with which each can exchange positions. Thus there are $2(n/2) = n$ swap neighbors $y \in N(x)$ that contain component $i \notin x$. Since there are $n^2/4$ total neighbors:

$$p_2 = \frac{n}{n^2/4} = \frac{\beta}{d}.$$

Again, these terms can be substituted into Equation (7) to yield the average objective function value over the swap neighborhood for the minimum graph bisection problem.

$$\text{avg } N(x) = f(x) + \frac{\alpha + \beta}{d}(\bar{f} - f(x)) = f(x) + \frac{2(n-1)}{n^2/4}(\bar{f} - f(x)).$$

This is the original difference equation with $k = \alpha + \beta = 2(n-1)$ and $d = n^2/4$.

We note here that it should be clear that the above analysis can be trivially extended to the (also NP-hard) maximum graph bisection problem. In this case, the objective is to maximize the weight in the cutset of (LHS, RHS).

4. Partial Neighborhoods on Elementary Landscapes

Suppose (X, N, f) is an elementary landscape. Recall from Definition 2 that a partial neighborhood of a neighborhood operator N is an operator N' such that, for all $x \in X$, $N'(x) \subseteq N(x)$. The partial neighborhoods we explore here have also the property that each component is exchanged out of x exactly α' times across the partial neighborhood where α' is fixed for all $x \in X$.

In this paper, we will observe two approaches to constructing partial neighborhoods. The first approach arises by exploiting structural regularity in the 2-opt neighborhood for TSP. In particular, we will see that the set of all 2-opt moves in which the length of the inverted tour segment is fixed yields the property that each edge is exchanged out α' times in the partial neighborhood. This observation allows one to compute the average objective function value over the partial neighborhood without explicitly evaluating any of the neighbors. We explore this approach in Section 4.1.
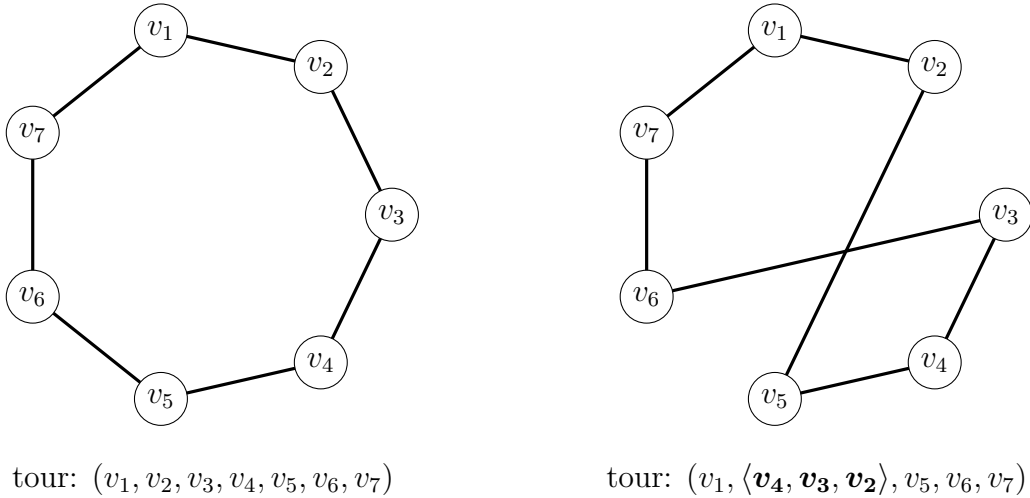
tour: $(v_1, v_2, v_3, v_4, v_5, v_6, v_7)$ tour: $(v_1, \langle \boldsymbol{v_4}, \boldsymbol{v_3}, \boldsymbol{v_2} \rangle, v_5, v_6, v_7)$

Figure 2: A 2-opt move that removes two edges and reconnects the tour inverts a subsequence of length ℓ in the vertex sequence corresponding to the tour. In this example, $\ell = 3$, and the inverted subsequence is highlighted in bold and placed in angle brackets.

The second approach for constructing partial neighborhoods is slightly more general, and arises by exploiting the fact that it is often possible to recognize a subset of the moves in a neighborhood that cannot yield an improving move. In this case, it is possible to focus on subsets of the neighborhood that only include moves that have the potential to contain an improving move. If the objective function measures some "error term" that contributes to $f(x)$ then we can consider moves that uniformly sample components that contribute to x, but only sample a subset of $C \setminus x$. We explore this approach for minimum graph bisection in Section 4.3 and weighted graph coloring in Section 4.2.

4.1. Partial Neighborhoods of the TSP

We begin by noting that every 2-opt move "reverses" a segment in a Hamiltonian tour.

Let $x \subseteq E$ be a set of edges that correspond to a Hamiltonian tour in the complete graph $G = (V, E)$. If we pick any vertex $v \in V$ as a starting vertex, the Hamiltonian tour described by x is given by an ordered sequence of vertices

$$(v = v_1, v_2, v_3, \ldots, v_n),$$

where each v_i is a distinct vertex in V and $(v_i, v_{i+1}) \in x$ for all $i \in \{1, 2, \ldots, n\}$. Here, and throughout, we assume the indexes are taken to be modulo n.

A 2-opt move that removes the edges (v, v_2) and $(v_{\ell+1}, v_{\ell+2})$ in x produces a neighbor $y = (x \setminus \{(v, v_2), (v_{\ell+1}, v_{\ell+2})\}) \cup \{(v, v_{\ell+1}), (v_2, v_{\ell+2})\}$. This corresponds to a new Hamiltonian tour:

$$(v = v_1, v_{\ell+1}, v_\ell, \ldots, v_2, v_{\ell+2}, v_{\ell+3}, \ldots, v_n).$$

Thus, the effect of the 2-opt operator is to reverse a subtour of ℓ vertices (see Figure 2).

Whitley and Ochoa [19] characterize how partial neighbors corresponding to 2-opt moves of length ℓ can be tracked by calculating changes in a partition matrix.

In the current paper we will do something much simpler, and potentially more useful. We will track partial neighborhoods corresponding to 2-opt moves of length ℓ where ℓ is bounded by a constant g. This has several advantages. First, it is cheaper to compute updates for partial neighborhoods when the segment being reversed is short. Second, many improving moves are often associated with 2-opt moves that reverse shorter segments. Third, the method used by Whitley and Ochoa looked at both edges that are removed and edges that are added into neighborhood structures as those neighborhoods are updated after a 2-opt move. But we already know that all of the edges that are removed are in x. Thus, we only need to track new edges that are introduced while constructing the neighborhood.

Consider the following illustration for all length 2 and 3 moves on a 7 vertex graph. We highlight a reversed segment in bold typeface and with angle brackets $\langle \rangle$.

length $\ell = 2$ moves							length $\ell = 3$ moves						
$\langle \boldsymbol{v_2}$	$\boldsymbol{v_1} \rangle$	v_3	v_4	v_5	v_6	v_7	$\langle \boldsymbol{v_3}$	$\boldsymbol{v_2}$	$\boldsymbol{v_1} \rangle$	v_4	v_5	v_6	v_7
$\langle \boldsymbol{v_3}$	$\boldsymbol{v_2} \rangle$	v_4	v_5	v_6	v_7	v_1	$\langle \boldsymbol{v_4}$	$\boldsymbol{v_3}$	$\boldsymbol{v_2} \rangle$	v_5	v_6	v_7	v_1
$\langle \boldsymbol{v_4}$	$\boldsymbol{v_3} \rangle$	v_5	v_6	v_7	v_1	v_2	$\langle \boldsymbol{v_5}$	$\boldsymbol{v_4}$	$\boldsymbol{v_3} \rangle$	v_6	v_7	v_1	v_2
$\langle \boldsymbol{v_5}$	$\boldsymbol{v_4} \rangle$	v_6	v_7	v_1	v_2	v_3	$\langle \boldsymbol{v_6}$	$\boldsymbol{v_5}$	$\boldsymbol{v_4} \rangle$	v_7	v_1	v_2	v_3
$\langle \boldsymbol{v_6}$	$\boldsymbol{v_5} \rangle$	v_7	v_1	v_2	v_3	v_4	$\langle \boldsymbol{v_7}$	$\boldsymbol{v_6}$	$\boldsymbol{v_5} \rangle$	v_1	v_2	v_3	v_4
$\langle \boldsymbol{v_7}$	$\boldsymbol{v_6} \rangle$	v_1	v_2	v_3	v_4	v_5	$\langle \boldsymbol{v_1}$	$\boldsymbol{v_7}$	$\boldsymbol{v_6} \rangle$	v_2	v_3	v_4	v_5
$\langle \boldsymbol{v_1}$	$\boldsymbol{v_7} \rangle$	v_2	v_3	v_4	v_5	v_6	$\langle \boldsymbol{v_2}$	$\boldsymbol{v_1}$	$\boldsymbol{v_7} \rangle$	v_3	v_4	v_5	v_6

All of the length $\ell = 2$ 2-opt moves form one partial neighborhood, and all of the length $\ell = 3$ moves form another partial neighborhood. While they

introduce different sets of new edges, they remove exactly the same set of edges: every edge in the solution x is removed twice.

Lemma 1. *Let x be a set of edges corresponding to a Hamiltonian tour. Let $(v_1, v_2, \ldots, v_n)$ with $v_i \in V$ for all $i \in \{1, 2, \ldots, n\}$ be some ordered sequence of vertices in the Hamiltonian tour defined by x. Let $N(x)$ be the standard 2-opt neighborhood of x. In other words, each edge set $y \in N(x)$ corresponds to a Hamiltonian tour that can be obtained by applying a 2-opt operation to x. Denote as $N^\ell(x) \subseteq N(x)$ the partial TSP neighborhood comprised of only those 2-opt moves that reverse a segment of length ℓ in the sequence of vertices. The average value of the neighbors in the partial neighborhood $N^\ell(x)$ is given by*

$$\operatorname{avg} N^\ell(x) = f(x) - \frac{2}{n} f(x) + \frac{2}{n} \sum_{i=1}^{n} w((v_i, v_{i+\ell})),$$

where the indexes are taken modulo n.

Proof. All partial neighborhoods constructed by taking only 2-opt moves of fixed length ℓ have n members because there are n unique reversible segments of length n over any Hamiltonian circuit. Thus $d' = n$.

Each edge in solution x is removed once at the beginning of a reversed segment and once at the end of a reversed segment. Thus, all the edges in x are removed twice, and $\alpha' = 2$.

For any 2-opt move that corresponds to the reversal of a segment of ℓ vertices, say from v_i to $v_{i+\ell-1}$, the edges

$$(v_{i-1}, v_i), (v_{i+\ell-1, i+\ell}) \in x$$

are removed, and the edges

$$(v_i, v_{i+\ell}), (v_{i-1}, v_{i+\ell-1}) \in C \setminus x$$

are added. Over all n positions in the ordered sequence, the contribution from the new edges are counted as follows.

$$\sum_{i=1}^{n} (w((v_i, v_{i+\ell})) + w((v_{i-1}, v_{i+\ell-1}))).$$

Since i is modulo n, shifting the second set of indices by one yields

$$\sum_{i=1}^{n} w((v_i, v_{i+\ell})) + w((v_{i-1}, v_{i+\ell-1})) = 2 \sum_{i=1}^{n} w(v_i, v_{i+\ell}).$$

Therefore,

$$\text{avg } N^{\ell}(x) = f(x) - \frac{2}{d'} f(x) + \frac{2}{d'} \sum_{i=1}^{n} w((v_i, v_{i+\ell})).$$

$\square$

When a local search algorithm computes a neighboring tour under 2-opt, it is not necessary to recompute the objective function value from scratch by summing all the edge weights in the new tour. It suffices to subtract the weights of the two edges removed, and add the weights of the two new edges. This obvious saving speeds up the objective function re-calculation by a linear factor. The following theorem proves that the average value corresponding to $N^{\ell}(x)$ for a partial 2-opt neighborhood can be also be quickly recomputed for a 2-opt neighbor y, i.e., $N^{\ell}(y)$.

Theorem 3. *Let g be a fixed constant. After a 2-opt move of any length $k < n/2$, updating the value $\text{avg } N^{\ell}(x)$ for all partial neighborhoods N^{ℓ} with $\ell \leq g < n/2$ can be done using only $O(g^2) = O(1)$ arithmetic operations.*

Proof. Again, suppose the ordered sequence of vertices $(v_1, v_2, \ldots, v_n)$ corresponds to the Hamiltonian tour given by the set of edges x. Let $P_{\ell}(x) = \sum_{i=1}^{n} w((v_i, v_{i+\ell}))$. Here, and throughout the remainder of the proof, we take the vertex indexes to be modulo n. Thus, by Lemma 1,

$$\text{avg } N^{\ell}(x) = f(x) - \frac{2}{n} f(x) + \frac{2}{n} P_{\ell}(x).$$

After a 2-opt move of length k from solution x to solution $y \in N(x)$, we need to calculate $P_{\ell}(y)$ from $P_{\ell}(x)$. It suffices to show that only a constant number of terms in the sum will change.

Denote as $(v_1', v_2', \ldots, v_n')$ the ordered sequence of vertices corresponding to the Hamiltonian tour given by the set of edges y. This is identical to the sequence corresponding to x up to an inversion of length $k < n/2$.

Suppose this inversion occurs between indexes a and b in $(v_1, v_2, \ldots, v_n)$. Without loss of generality, we suppose $1 < a < b < n$. Note that $b - a = k - 1$.

$$\cdots \quad v_{a-2} \quad v_{a-1} \;\vdots\; v_a \quad v_{a+1} \quad v_{a+2} \quad \cdots \quad v_{b-2} \quad v_{b-1} \;\; v_b \;\vdots\; v_{b+1} \quad v_{b+2} \quad \cdots$$

$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad}_{\text{length-}k\text{ segment to reverse}}$$
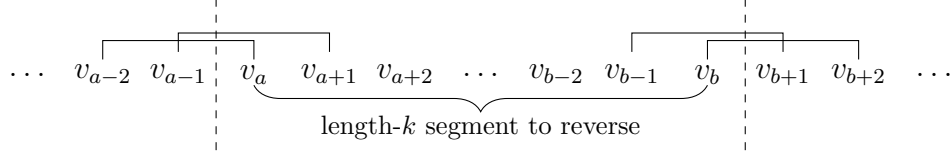
Figure 3: A vertex sequence corresponding to a Hamiltonian tour x showing which edges must be updated to compute $P_\ell(y)$ from $P_\ell(x)$ (for $\ell = 2$) after a length $k < n/2$ 2-opt move that reverses the subsequence from v_a to v_b. In particular, the set of edges that must be updated are $\{(v_{a-2}, v_a), (v_{a-1}, v_{a+1}), (v_{b-1}, v_{b+1}), (v_b, v_{b+2})\}$.

To construct y from x, we must remove edges (v_{a-1}, v_a) and (v_b, v_{b+1}) and insert edges (v_{a-1}, v_b) and (v_a, v_{b+1}). This means that for all $1 \le i < a$, $(v_{i-1}, v_i) \in x \wedge (v_{i-1}, v_i) \in y$, and for all $b < i \le n$, $(v_i, v_{i+1}) \in x \wedge (v_i, v_{i+1}) \in y$.

If $b < i \le n$, then all edges $(v_i, v_{i+\ell})$ contribute to both $P_\ell(x)$ and $P_\ell(y)$. Similarly, if $1 \le i < a - \ell$, then all edges $(v_i, v_{i+\ell})$ contribute to both $P_\ell(x)$ and $P_\ell(y)$. The contribution from edges $(v_i, v_{i+\ell})$ where $a - \ell \le i < a$ or $b - \ell + 1 \le i \le b$ must be updated. However, there are at most 2ℓ such edges.

The remaining edges correspond to the set of all edges $(v_i, v_{i+\ell})$ where $a \le i < b - \ell + 1$. Note that if such an edge $(v_i, v_{i+\ell})$ contributed to $P_\ell(x)$, there exists a v'_j in the ordered sequence of vertices corresponding to y such that $v_{i+\ell} = v'_j$. In this case, by symmetry, $v_i = v'_{j+\ell}$. Therefore, the reversal does not change the set of edges that contribute to both $P_\ell(x)$ and $P_\ell(y)$ for these edges. Figure 3 gives an illustration of the intervals of length $\ell = 2$ that need to be updated after such a move.

Therefore the updates to the summation terms only need to occur for the indexes in the ranges $a - \ell$ to $a - 1$ and $b - \ell + 1$ to b as follows.

$$P_\ell(y) = P_\ell(x) + \left(\sum_{i=a-\ell}^{a-1} w((v'_i, v'_{i+\ell})) - w((v_i, v_{i+\ell})) \right)$$
$$+ \left(\sum_{i=b-\ell+1}^{b} w((v'_i, v'_{i+\ell})) - w((v_i, v_{i+\ell})) \right).$$

Using a partial update scheme, $f(y)$ can be computed from $f(x)$ by subtracting the contribution of the two edges removed and adding the contribution of the two edges added. This requires only a constant number of operations. Similarly, the number of arithmetic operations needed to compute $P_\ell(y)$ from
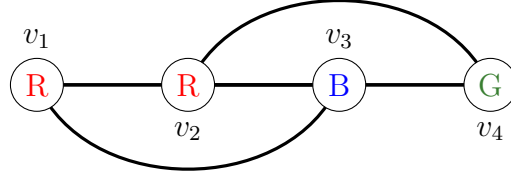
23

Figure 4: A graph coloring problem with vertices $V = \{v_1, v_2, v_3, v_4\}$. Invisible edges have weight zero.

$P_\ell(x)$ for all $2 \le \ell \le g$ is bounded by $\sum_{i=2}^{g} 2i = O(g^2)$, which is $O(1)$ relative to n. $\qquad\square$

4.2. Partial Neighborhoods for Graph Coloring

For graph coloring we explore a slightly different question with regard to partial neighborhoods. When searching the graph coloring neighborhood, certain moves do not appear to be reasonable if we are interested in removing conflicts (we will later show that the same is true for the minimum graph bisection problem).

A weighted graph coloring problem is given in Figure 4 where $V = \{v_1, v_2, v_3, v_4\}$. All edges have weight zero except for the set

$$\{(v_1, v_2), (v_2, v_3), (v_3, v_4), (v_1, v_3), (v_2, v_4)\}.$$

The set of colors is $\{R, G, B\}$. In Figure 4, vertices v_1 and v_2 are the same color R, v_3 is color B, and v_4 is color G. The only nonzero weight edge in conflict is (v_1, v_2). Changing the color of v_1 and v_2 can remove the conflict, so recoloring v_1 and v_2 are reasonable neighborhood moves. But changing the color of v_3 and v_4 is useless: recoloring these vertices cannot remove existing conflicts, but in fact can generate new conflicts.

Whitley and Sutton [21] calculated the average of a partial neighborhood for the graph coloring problem where the moves are restricted to those that remove conflicts that contribute to the objective function. We present a new theorem that generalizes their results to weighted graph coloring and holds for a larger class of partial neighborhoods.

Let $G = (V, E)$ be the complete graph over $|V| = n$ vertices. As in Section 3.2, let $C = E$ be the set of components (edges) and let $w : C \to \mathbb{R}_{\ge 0}$ be the nonnegative weight function over the components. We define a map

$\sigma : V \to \mathbb{R}_{\geq 0}$ as follows

$$\sigma(v) = \sum_{(u,v) \in E} w((u,v)).$$

Thus $\sigma(v)$ is the sum of weights of all edges incident to v. We also define the map $S : 2^V \to \mathbb{R}_{\geq 0}$ as follows. Let $Q \subseteq V$

$$S(Q) = \sum_{v \in Q} \sigma(v).$$

Note that

$$S(V) = \sum_{v \in V} \sigma(v) = 2 \sum_{i \in C} w(i),$$

since the weight of each edge $i = (u,v)$ is counted twice: once in $\sigma(u)$ and once again in $\sigma(v)$.

Let $Q_x \subseteq V$ be defined as

$$Q_x = \{v \in V \mid \exists (u,v) \in x\},$$

that is, the set of vertices that appear in edges in the conflict set x.

Definition 3. *Let X be the collection of all conflict sets corresponding to r-colorings on V. Let N be the recoloring neighborhood defined in Section 3.2 and let Q be any set $Q_x \subseteq Q \subseteq V$. A partial recoloring neighborhood based on Q, N^Q, is defined as follows. For every $x \in X$, $N^Q(x) \subseteq N(x)$ is a subset that contains elements of $N(x)$ such that only the vertices in Q are recolored.*

Theorem 4. *Let N^Q be the partial recoloring neighborhood from Definition 3. The neighborhood average over $N^Q(x)$ on the weighted graph coloring problem is given by:*

$$\text{avg } N^Q(x) = f(x) + \frac{S(Q) - 2rf(x)}{|Q|(r-1)}.$$

Proof. Only vertices in Q will be recolored, and the size of the neighborhood is $d' = |Q|(r-1)$. But the number of ways a vertex can remove conflicts does not change, thus:

$$p_1' = \frac{2(r-1)}{|Q|(r-1)}.$$

The edges that are not in the conflict set x that might be affected when vertices in Q are recolored are of the following two types.

1. Edges incident to one vertex in Q.
2. Edges which are incident to two vertices in Q.

$S(Q)$ sums the edge weights associated with all of the vertices that are being recolored. Consider vertices u and v. If only one of the vertices is in Q, then $w((u,v))$ contributes only once to $S(Q)$. If both $u, v \in Q$, then $w((u,v))$ contributes twice to $S(Q)$.

If an edge appears in x, its associated weight is also counted twice in $S(Q)$ since $Q \supseteq Q_x$. Therefore, the sum of weights of edges in graph G that do *not* appear in x but appear in neighbors when the vertices in Q are recolored is given by $S(Q) - 2f(x)$.

Consider an edge where one vertex is in Q and one vertex is not. Obviously, the edge does not appear in x since $Q \supseteq Q_x$. There is only one way to generate a conflict, so $\beta' = 1$.

$$p_2' = \frac{1}{|Q|(r-1)}.$$

Finally, consider an edge (u,v) that does not appear in x, but where $u, v \in Q$. In this case we have included the weight associated with (u,v) twice in the quantity $S(Q) - 2f(x)$. But this edge will appear in two neighbors in the partial neighborhood: once when u is recolored to the color of v, and once when v is recolored the color of u. The total weight contribution of edges not in x to the average objective function value of the neighbors is

$$\frac{1}{d'}(S(Q) - 2f(x)) = p_2'(S(Q) - 2f(x)).$$

We can now substitute these terms into the average neighborhood equation for the component model

$$\text{avg } N^Q(x) = f(x) - p_1'f(x) + p_2'(S(Q) - 2f(x))$$
$$\text{avg } N^Q(x) = f(x) - \frac{2(r-1)}{d'}f(x) + \frac{1}{d'}(S(Q) - 2f(x))$$
$$= f(x) + \frac{S(Q) - 2rf(x)}{|Q|(r-1)}.$$

$\square$

This partial neighborhood is not elementary because the above difference equation depends locally on x and Q. Nevertheless, the average value of all of the neighbors in the dynamically defined neighborhood can be cheaply computed exploiting a decomposition of the recoloring neighborhood introduced in Section 3.2. However, since Q can be any subset of V that also includes Q_x, let $Q = V$, so that $S(Q) = 2 \sum_{i \in C} w(i)$ and we obtain:

$$
\begin{aligned}
\operatorname{avg} N^Q(x) &= f(x) + \frac{S(Q) - 2rf(x)}{|Q|(r-1)} \\
&= f(x) + \frac{2 \left(\sum_{i \in C} w(i) \right) - 2rf(x)}{|V|(r-1)} \\
&= f(x) + \frac{2r\bar{f} - 2rf(x)}{|V|(r-1)} \\
&= f(x) + \frac{2r}{|V|(r-1)}(\bar{f} - f(x)).
\end{aligned}
$$

which yields the original wave equation.

We can also relate the average of the partial neighborhood to the average of the full neighborhood.

$$
\operatorname{avg} N^Q(x) \leq \operatorname{avg} N(x) \iff \frac{S(Q) - 2rf(x)}{|Q|(r-1)} \leq \frac{2r\bar{f} - 2rf(x)}{|V|(r-1)}.
$$

Also note that

$$
\operatorname{avg} N^Q(x) < f(x) \iff S(Q) < 2rf(x) \iff \frac{S(Q)}{2r} < f(x),
$$

and thus under these conditions an improving move is guaranteed to exist. Since $2r$ is a constant, and $f(x)$ is fixed by x we minimize the value of $\operatorname{avg} N'(x)$ by keeping $S(Q)$ minimal. Assume that $S(Q)$ is minimal. When $f(x) > \bar{f}$, we already know that all full neighborhoods include an improving move. When $f(x) < \bar{f}$, if r is sufficiently large that all partial neighborhoods include an improving move, then the search space becomes unimodal.

Empirically, we have observed cases where

$$
\operatorname{avg} N^Q(x) < f(x) < \operatorname{avg} N(x) < \bar{f},
$$

where the global optimum was contained in the partial neighborhood $N^Q(x)$. Thus, calculations about partial neighborhoods can point to improving moves even when $f(x)$ is already a near-optimal (but not locally optimal) solution.

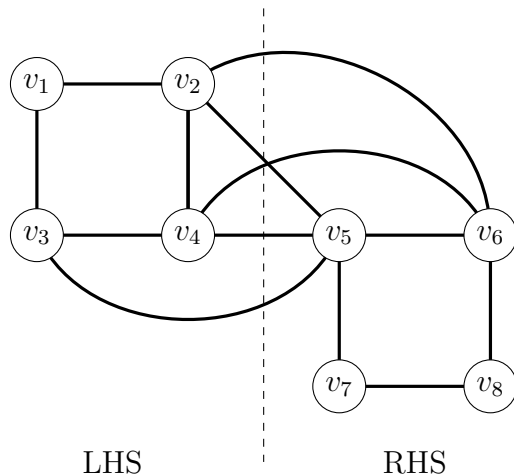As $f(x)$ gets smaller, there are fewer conflicts and it is cheaper to compute $S(Q)$.

Figure 5: An instance of the minimum graph bisection problem with LHS = $\{v_1, v_2, v_3, v_4\}$ and RHS = $\{v_5, v_6, v_7, v_8\}$. Invisible edges have weight zero.

4.3. Partial Neighborhoods for Minimum Graph Bisection

Figure 5 shows a minimum graph bisection problem with 8 vertices and a partition where LHS = $\{v_1, v_2, v_3, v_4\}$ and RHS = $\{v_5, v_6, v_7, v_8\}$. In this graph, edges with zero weight are invisible.

As is the case with graph coloring, there are moves that are included in the swap neighborhood for the minimum graph bisection problem that also are ineffective. Without loss of generality, suppose a vertex u is in the LHS partition, and all edges that connect it to vertices in the RHS partition have weight zero. Furthermore, suppose that a vertex v is in the RHS partition, and all edges that connect it to vertices in the LHS partition have weight zero. In the swap neighborhood operator, it is unreasonable to swap the pair u and v because this cannot reduce the objective function relative to the current solution x.

We again define a map $\sigma : V \to \mathbb{R}_{\geq 0}$ as follows. For all $v \in V$,

$$\sigma(v) = \sum_{u \in V} w((u, v)).$$

Thus $\sigma(v)$ is the sum of all weights on the edges incident to v.

Similar to the graph coloring problem, we define the set $Q_x \subseteq V$ as

$$Q_x = \{v \in V \mid \exists (u, v) \in x \text{ such that } w((u, v)) \neq 0\}.$$

28

This is the set of all vertices appearing in an edge in the cutset of $(\mathrm{LHS}, \mathrm{RHS})$ that has nonzero weight under w.

Definition 4. *Let N be the swap neighborhood defined in Section 3.3 and let Q be any set $Q_x \subseteq Q \subseteq V$. A* partial swap neighborhood *based on Q, N^Q, is defined as follows. For every $x \in X$, $N^Q(x) \subseteq N(x)$ is a subset that contains elements of $N(x)$ such that only the vertices in Q are swapped.*

Let $Q_L = Q \cap \mathrm{LHS}$ and $Q_R = Q \cap \mathrm{RHS}$ partition Q. The neighborhood size $d' = |N^Q(x)|$ of the partial swap neighborhood is clearly $|Q_L||Q_R|$.

In case $|Q_R| = |Q_L|$ then the vertices in the LHS and RHS would be acted on uniformly; and we could force this to be the case by enlarging the set of vertices in Q. Instead, we will be more general and define two functions $S_L, S_R : 2^V \to \mathbb{R}_{\geq 0}$ as follows

$$S_L(Q) = \sum_{v \in Q_L} \sigma(v) \qquad \text{and} \qquad S_R(Q) = \sum_{v \in Q_R} \sigma(v).$$

Theorem 5. *Let N^Q be a partial swap neighborhood defined in Definition 4. The neighborhood average for $N^Q(x)$ on the minimum graph bisection problem is given by:*

$$\mathrm{avg}\, N^Q(x) = f(x) + \frac{2(1 - |Q_R| - |Q_L|)f(x) + |Q_L|S_R(Q) + |Q_R|S_L(Q)}{|Q_R||Q_L|}.$$

Proof. Assume $(u, v) \in x$ such that $u \in Q_L$ and $v \in Q_R$ and $w((u, v))$ is nonzero. Clearly, swapping u and v does not remove (u, v) from x. To remove (u, v) from x, u can be swapped with any vertex in $Q_R \setminus \{v\}$ (and there are $|Q_R| - 1$ such swaps), or v can be swapped with any vertex in $Q_L \setminus \{u\}$ (and there are $|Q_L| - 1$ such swaps).

Therefore, the number of neighbors $y \in N^Q(x)$ in which $(u, v) \notin y$ is $\alpha' = (|Q_L| - 1) + (|Q_R| - 1)$. Thus we have

$$p_1' = \alpha'/d' = \frac{|Q_R| - 1 + |Q_L| - 1}{|Q_R||Q_L|}.$$

We next count the number of times an edge that does not appear in x appears in a neighbor $y \in N^Q(x)$. Note that the sum $S_L(Q) + S_R(Q)$ double counts the weights of edges with both endpoints in Q, but only counts once the weight of each edge with only one endpoint in Q.

Since $Q_x \subseteq Q$, each edge in x with a nonzero weight has one vertex in Q_L and one vertex in Q_R so its weight is counted once in $S_L(Q)$ and once in $S_R(Q)$. We can isolate the weights of edges that do not appear in x by subtracting $f(x)$ from both $S_L(Q)$ and $S_R(Q)$. Vertices in Q_L move to the RHS a total of $|Q_R|$ times. Vertices in Q_R move to the LHS a total of $|Q_L|$ times. Therefore an edge $(u, v) \notin x$ with $u \in Q_L$ will appear in $|Q_R|$ distinct neighbors $y \in N^Q(x)$. The total weight contribution of such edges to the neighbors is hence $|Q_R| (S_L(Q) - f(x))$. Similarly, an edge $(u, v) \notin x$ with $u \in Q_R$ will appear in $|Q_L|$ distinct neighbors $y \in N^Q(x)$. The total weight contribution of these edges to the neighbors is $|Q_L| (S_R(Q) - f(x))$. Therefore,

$$p_2' = \frac{\beta'}{|Q_R||Q_L|} = \frac{(S_L(Q) - f(x)) |Q_R| + (S_R(Q) - f(x))|Q_L|}{|Q_R||Q_L|}.$$

Substituting these terms into the average value equation yields

$$\text{avg } N^Q(x) = f(x) - \frac{|Q_R| - 1 + |Q_L| - 1}{|Q_R||Q_L|} f(x) + \frac{|Q_L|(S_R(Q) - f(x)) + |Q_R|(S_L(Q) - f(x))}{|Q_R||Q_L|}$$

$$= f(x) + \frac{2(1 - |Q_R| - |Q_L|)f(x) + |Q_L|S_R(Q) + |Q_R|S_L(Q)}{|Q_R||Q_L|}.$$

$\square$

Again, this partial neighborhood is not elementary because the above difference equation depends locally on x. However, when $Q = V$, we have

$$S_L(V) + S_R(V) = 2 \sum_{i \in C} w(i) = 2 \left(\frac{2n - 2}{n} \bar{f} \right),$$

and $|Q_R| = |Q_L| = n/2$. We can recover the original wave equation as follows

$$\text{avg } N^Q(x) = f(x) + \frac{2f(x)(1 - |Q_R| - |Q_L|) + |Q_L|S_R(Q) + |Q_R|S_L(Q)}{|Q_R||Q_L|}$$

$$= f(x) + \frac{2f(x)(1 - n/2 - n/2) + (n/2)(S_R(Q) + S_L(Q))}{n^2/4}$$

$$= f(x) + \frac{2(1 - n)f(x) + n/2(2\frac{2n-2}{n}\bar{f})}{n^2/4} = f(x) + \frac{2(1-n)f(x) + 2(n-1)\bar{f}}{n^2/4}$$

$$= f(x) + \frac{2(n-1)}{n^2/4}(\bar{f} - f(x)).$$

By comparing the total weight of the components removed from x to the weight of the components that do not appear in x but appear in neighbors of x, we can see that

$$\text{avg } N^Q(x) < f(x) \iff (|Q_L|S_R(Q) + |Q_R|S_L(Q)) < 2(|Q_R| + |Q_L| - 1)f(x),$$

and

$$\text{avg } N^Q(x) < f(x) \iff \frac{|Q_L|S_R(Q) + |Q_R|S_L(Q)}{2(|Q_R| + |Q_L| - 1)} < f(x),$$

and thus under these conditions an improving move is guaranteed to exist.

When using this information in a concrete search algorithm, one is interested in the computational efficiency of calculating these quantities. The map σ can be stored as an n-element array and computed only once. If Q_L and Q_R are chosen to minimal in size, then $Q = Q_x$ and every vertex in the two subsets will be associated with an edge $(u, v) \in x$. Therefore, when $f(x)$ is calculated, we can also construct a list of which vertices contribute to $S_L(Q_x)$ or $S_R(Q_x)$. Furthermore, if a move from x to $y \in N^Q(x)$ results in the calculation of $f(y)$, we can also compute a partial update to the list of vertices contributing to $S_L(Q_y)$ or $S_R(Q_y)$ so that these sums can be incrementally updated during a move.

5. Conclusions

This paper has employed a component based model to reconstruct elementary landscape proofs for a number of combinatorial optimization problems. Furthermore, this paper has examined ways in which partial neighborhoods can be evaluated by exploiting knowledge about the elementary landscape structures of the traveling salesperson problem, graph coloring problem, and minimum graph bisection problem. These results are important because neighborhoods that restrict search to the most promising moves are more likely to be used by modern local search algorithms.

But more than this, statistical information about the expected value of points in the search landscape that is one or more moves ahead can potentially be useful in guiding search. This is a new, still relatively unexplored idea. Exploiting statistical information about partial neighborhoods is also a relatively new idea. The current paper provides a foundation for future applications exploiting statistical information about neighborhood and partial neighborhoods of elementary landscapes.

Acknowledgements

[1] Preston Briggs. *Register Allocation via Graph Coloring.* PhD thesis, Rice University, Houston, Texas, USA, 1992.

[2] Francisco Chicano, L. Darrell Whitley, Enrique Alba, and Francisco Luna. Elementary landscape decomposition of the frequency assignment problem. *Theoretical Computer Science*, 412(43):6002–6019, 2011.

[3] Bruno Codenotti and Luciano Margara. Local properties of some NP-complete problems. Technical Report TR 92-021, International Computer Science Institute, Berkeley, CA, 1992.

[4] Bruce Colletti and J. Wesley Barnes. Linearity in the traveling salesman problem. *Applied Mathematics Letters*, 13(3):27–32, April 2000.

[5] Boryana Dimova, J. Wesley Barnes, and Elmira Popova. Arbitrary elementary landscapes & AR(1) processes. *Applied Mathematics Letters*, 18(3):287–292, 2005.

[6] Lov K. Grover. Local search and the local structure of NP-complete problems. *Operations Research Letters*, 12:235–243, 1992.

[7] Shen Lin and Brian W. Kernighan. An effective heuristic algorithm for the traveling salesman problem. *Operations Research*, 21:498–516, 1973.

[8] Samadhi Nallaperuma, Andrew M. Sutton, and Frank Neumann. Fixed-parameter evolutionary algorithms for the Euclidean traveling salesperson problem. In *Evolutionary Computation (CEC), 2013 IEEE Congress on*, pages 2037–2044, 2013.

[9] Christos H. Papadimitriou and Kenneth Steiglitz. *Combinatorial optimization: algorithms and complexity.* Courier Dover Publications, 1998.

[10] Rong Qu, Edmund K. Burke, Barry McCollum, Liam T. G. Merlot, and S. Y. Lee. A survey of search methodologies and automated system development for examination timetabling. *J. Scheduling*, 12(1):55–89, 2009.

[11] Jonathan E. Rowe and Michael D. Vose. Unbiased black box search algorithms. In Natalio Krasnogor and Pier Luca Lanzi, editors, *GECCO*, pages 2035–2042. ACM, 2011.

[12] Andrew Solomon, J. Wesley Barnes, Steftcho P. Dokov, and Raul Acevedo. Weakly symmetric graphs, elementary landscapes, and the TSP. *Applied Mathematics Letters*, 16(3):401–407, 2003.

[13] Peter F. Stadler. Toward a theory of landscapes. In Ramon Lopéz-Peña, Riccardo Capovilla, Ricardo García-Pelayo, Henri Waelbroeck, and Federico Zertruche, editors, *Complex Systems and Binary Networks*, pages 77–163. Springer Verlag, 1995.

[14] Peter F. Stadler. Landscapes and their correlation functions. *Journal of Mathematical Chemistry*, 20:1–45, 1996.

[15] Peter F Stadler and Wolfgang Schnabl. The landscape of the traveling salesman problem. *Physics Letters A*, 161(4):337–344, 1992.

[16] Andrew M. Sutton, L. Darrell Whitley, and Adele E. Howe. Computing the moments of k-bounded pseudo-Boolean functions over Hamming spheres of arbitrary radius in polynomial time. *Theoretical Computer Science*, 425:58–74, 2012.

[17] Darrell Whitley and Wenxiang Chen. Constant time steepest descent local search with lookahead for NK-landscapes and MAX-kSAT. In Terence Soule and Jason H. Moore, editors, *GECCO*, pages 1357–1364. ACM, 2012.

[18] Darrell Whitley, Wenxiang Chen, and Adele E. Howe. An empirical evaluation of $O(1)$ steepest descent for NK-landscapes. In Carlos A. Coello Coello, Vincenzo Cutello, Kalyanmoy Deb, Stephanie Forrest, Giuseppe Nicosia, and Mario Pavone, editors, *PPSN (1)*, volume 7491 of *Lecture Notes in Computer Science*, pages 92–101. Springer, 2012.

[19] Darrell Whitley and Gabriela Ochoa. Partial neighborhoods of the traveling salesman problem. In Natalio Krasnogor and Pier Luca Lanzi, editors, *GECCO*, pages 529–536. ACM, 2011.

[20] L. Darrell Whitley, Francisco Chicano, Enrique Alba, and Francisco Luna. Elementary landscapes of frequency assignment problems. In

Martin Pelikan and Jürgen Branke, editors, *GECCO*, pages 1409–1416. ACM, 2010.

[21] L. Darrell Whitley and Andrew M. Sutton. Partial neighborhoods of elementary landscapes. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-09)*, 2009.