Correlation Structure of the Landscape of the Graph-Bipartitioning-Problem

By

PETER F. STADLER[a,b,†] AND ROBERT HAPPEL[b]

[a] Max Planck Institut für Biophysikalische Chemie

(Karl Friedrich Bonhoeffer Institut)

Dept. 080 – Biochemische Kinetik

Am Fassberg, D-3400 Göttingen, Germany

[b]Institut für Theoretische Chemie

Universität Wien

Währingerstraße 17, A-1090 Vienna, Austria

Phys. Abstr. Classification: 0250, 0550

[†]Correspondence to:

Peter F. Stadler

Institut für Theoretische Chemie, Universität Wien

Währingerstraße 17, A-1090 Wien, Austria

Abstract

The relaxation of random walks and the autocorrelation function on the landscape of the graph-bipartitioning problem (GBP) are calculated.

1. Introduction

The investigation of the correlation structure of combinatorial optimization problems has gained considerable interest recently.

One of the combinatorial optimization problems which have been studied in detail is the **Graph Bipartitioning Problem** (GBP); *cf.* Fu and Anderson (1986), Liao (1987), Banavar *et al.* (1987), Wiethege and Sherrington (1987) or Fu (1989). Given a graph with an even number n of vertices and an associated matrix H, the task is to find a partition of the vertex set V into two equal-sized subsets A and B such that

$$f([A;B]) = \sum_{i\in A}\sum_{j\in B} h_{ij} \tag{1}$$

is minimized. In a variant of the problem the graph is edge weighted with weights h_{ij} along an edge connecting the vertices i and j. We will treat (as usual) the special case where the h_{ij} , $i \le j$, are mutually independent random variables.

The GBP is closely related to the Sherrington-Kirkpatrick Spin-glass. In fact, the cost function may be viewed as an SK-Hamlitonian with the constraint of vanishing total spin, and the expected value of the global optimum can also be related to the ground state of the SK model. (Fu and Anderson 1986).

The main difference of the two models from the point of view of optimization heuristics consists of an entirely different topology of the configuration space $\mathcal{C}$. For the SK model, the canonical metric is induced by single-spin flips. This is tantamount to the Hamming metric and thus $\mathcal{C}_{SK} = \mathbf{B}^n$, a binary hypercube of dimension n.

The configuration space $\mathcal{C}$ of the GBP consists of the set of all partitions of the vertex set in two equally sized subsets A and B. A configuration can be encoded as a binary string by labelling a vertex by "1" if it contained in subset A and by "0" if it is contained in B. The canonical move set consists of exchanges of single vertices, defining two equally sized partitions of V as neighbours of each other iff the symmetric differences[1]

$$A \ominus C = D \ominus B = \{v_1, v_2\} \tag{2}$$

both equal the pair of exchanged vertices. With this definition of neighborhood the set $\mathcal{C}$ can be viewed as a graph. The minimum number of pair exchanges $d([A;B],[C;D])$ necessary to convert $[A;B]$ into $[C;D]$ is then the minimal number of edges separating the vertices $[A;B]$ and $[C;D]$ in the "configuration graph" $\mathcal{C}$ and is thus a metric.

Let $\$p$ denote the encoding of a partition p as a binary string. It is easily seen that the usual distance between vertices of a graph, $d_\Gamma([A;B],[C;D])$ and the Hamming distance $d_H(\$[A,B],\$[C,D])$ are related by

$$d_\Gamma(p_1,p_2) = \frac{1}{2} d_H(\$p_1,\$p_2) \tag{3}$$

[1] The symmetric difference of two sets X and Y is defined as

$X \ominus Y = (X \setminus Y) \cup (Y \setminus X) = (X \cup Y) \setminus (X \cap Y)$.

The distance sequence, **ds**, i.e. the number of configurations with a given distance from an arbitrary reference point, the number of configurations, $\#\mathcal{C}$, and the diameter in the configuration space, diam$\mathcal{C}$, are easily obtained:

$$\begin{aligned}
\mathbf{ds}\{(\mathcal{C}, d_\Gamma)\} &= \Big\{ \binom{n/2}{d}^2 \Big\} \\
\#\mathcal{C} &= \binom{n}{n/2} \\
\mathrm{diam}\ (\mathcal{C}, d_\Gamma) &= n/2
\end{aligned} \tag{4}$$

By using the string representation, $\mathcal{C}$ is imbedded in the Boolean hypercube $\mathbf{B}^n$. However

$$\frac{\#\mathcal{C}}{\#\mathbf{B}^n} = 2^{-n} \cdot \binom{n}{n/2} \sim \frac{1}{\sqrt{2\pi n}} \to 0 \tag{5}$$

thus for large problems most vertices of the Boolean hypercube do not code for valid configurations of the GBP and the GBP is a quite strongly constrained relative of the SK-spin glass.

2. Relaxation of Random Walks

The statistical properties of random walks on the graph $\mathcal{C}$ are completely contained in the probabilities ϕ_{sd} for a random walk to be in a distance d from the starting point after s steps. In general this probability distribution fulfils the following recursion relations on any distance transitive graph:

$$\begin{aligned}
&\phi_{sd} = a^+_{d-1} \cdot \phi_{s-1,d-1} + a^0_d \cdot \phi_{s-1,d} + a^-_{d+1} \cdot \phi_{s-1,d+1} \\
&\phi_{00} = 1 \\
&\phi_{sd} = 0 \quad \text{for } d > s
\end{aligned} \tag{6}$$

The coefficients a^+_d, a^0_d and a^-_d denote the probability of making a step "forward", "sideward" and "backward", respectively, given the walk is in a distance d from the starting point. For the GBP-graph $\mathcal{C}$ one obtains easily

$$\begin{aligned}
a^+_d &= (n-2d)^2/n^2 \\
a^0_d &= 4d(n-2d)/n^2 \\
a^-_d &= 4d^2/n^2
\end{aligned} \tag{7}$$

Although we have no closed solution for ϕ_{sd}, we can obtain some insight into the relaxation behaviour of random walks from the expected values of the distance and the squared distance from the starting point after s steps along the walk

$$\begin{aligned}
\Delta_1(s) &= \sum_{d=0}^{s} d \cdot \phi_{sd} \\
\Delta_2(s) &= \sum_{d=0}^{s} d^2 \cdot \phi_{sd}
\end{aligned} \tag{8}$$

Using the recursion relation for ϕ_{sd} one obtains after some simple algebra a system of linear inhomogeneous difference equations

$$\begin{aligned}\Delta_1(s) &= (1-\frac{4}{n})\cdot\Delta_1(s-1)+1\\ \Delta_2(s) &= (1-\frac{8}{n}+\frac{8}{n^2})\cdot\Delta_2(s-1)+(2-\frac{4}{n})\cdot\Delta_1(s-1)+1\end{aligned}\tag{9}$$

The fixed points of this difference equations are unique and correspond to the limit $s\to\infty$, or equivalently, random sampling.

$$\begin{aligned}\Delta_1(\infty) &= \langle d(x,y)\rangle_{random} = \frac{n}{4}\\ \Delta_2(\infty) &= \langle d^2(x,y)\rangle_{random} = \frac{n^2}{16}\cdot\frac{n}{n-1}\end{aligned}\tag{10}$$

It is convenient to introduce the corresponding *relaxation functions*

$$q_k(s) = \frac{\langle d^k(x,y)\rangle_{random} - \langle d^k(x_0,x_s)\rangle}{\langle d^k(x,y)\rangle_{random}} = 1-\frac{\Delta_k(s)}{\Delta_k(\infty)} \quad k=1,2 \tag{11}$$

The difference equations for $q_k(s)$ read

$$\begin{aligned}q_1(s) &= (1-\frac{4}{n})q_1(s-1)\\ q_2(s) &= (1-\frac{8}{n}+\frac{8}{n^2})q_2(s-1)+\frac{8(n-1)(n-2)}{n^3}q_1(s-1)\end{aligned}\tag{12}$$

For $q_1(s)$ we obtain immediately

$$q_1(s) = (1-\frac{4}{n})^s = e^{-s/\tau_1} \tag{13a}$$

and for $q_2(s)$ an ansatz with two different relaxation rates yields readily

$$\begin{aligned}q_2(t) &= 2\frac{n-1}{n}(1-\frac{4}{n})^t - \frac{n-2}{n}(1-\frac{8}{n}+\frac{8}{n^2})^t =\\ &= 2\frac{n-1}{n}\cdot e^{-t/\tau_1} - \frac{n-2}{n}\cdot e^{-t/\tau_2}\end{aligned}\tag{13b}$$

The relaxation times are

$$\begin{aligned}\tau_1 &= \frac{1}{4}(n+2)+\mathcal{O}(\frac{1}{n})\\ \tau_2 &= \frac{1}{8}(n-3)+\mathcal{O}(\frac{1}{n})\end{aligned}\tag{14}$$

The variance of ϕ_{sd} as a function of the number of steps is of course given by

$$\sigma^2(s) = \Delta_2(s) - \Delta_1^2(s) \tag{15}$$

3. Autocorrelation Functions

Weinberger (1990) proposed the autocorrelation function

$$\rho(d) = \frac{\langle (f(x)-\bar{f})(f(y)-\bar{f})\rangle_{d(x,y)=d}}{\mathrm{var}\ f} \tag{16}$$

as the most useful characteristic of a fitness landscape $f:\mathcal{C}\to\mathbb{R}$. Apart from totally uncorrelated landscapes, $\rho(d)=\delta(d,0)$, the simplest class consists of the nearly fractal AR(1) landscapes (Sorkin 1988, Weinberger and Stadler 1992) characterized by

$$\rho(d) \approx \exp -\frac{d}{\lambda}, \qquad d \ll n \tag{17}$$

The definition of the autocorrelation function, equ.(16) can be rewritten as

$$\rho(d) = 1 - \frac{\langle (f(x)-f(y))^2\rangle_{d(x,y)=d}}{\langle (f(x)-f(y))^2\rangle_{random}} \tag{18}$$

Now let us consider two partitions $[A,B]$ and $[C,D]$. Let

$$\begin{aligned} S &= A\cap C \qquad T = B\cap D \\ P &= A\cap D \qquad Q = B\cap C \end{aligned} \tag{19}$$

then the difference of the cost functions becomes

$$\begin{aligned} & f([A;B]) - f([C;D]) = \\ & \sum_{i\in S,j\in T} + \sum_{i\in P,j\in T} + \sum_{i\in S,j\in Q} + \sum_{i\in P,j\in Q} - \sum_{i\in S,j\in T} - \sum_{i\in Q,j\in T} - \sum_{i\in S,j\in P} - \sum_{i\in Q,j\in P} \\ & \quad = \sum_{i\in P}\sum_{j\in T} h_{ij} - \sum_{i\in Q}\sum_{j\in T} h_{ij} + \sum_{i\in S}\sum_{j\in Q} h_{ij} - \sum_{i\in S}\sum_{j\in P} h_{ij} \end{aligned} \tag{20}$$

Observe that $d = \#P = \#Q$ and $n/2 - d = \#S = \#T$. Thus we have for the expected value of the squared distance (using that all sums consist of different independent random variables and are thus again mutually independent)

$$\begin{aligned}
&\langle (f([A;B]) - f([C;D]))^2 \rangle = \\
&\qquad = 2\Big\langle \big(\sum_{i\in P}\sum_{j\in T} h_{ij} - \sum_{i\in Q}\sum_{j\in T} h_{ij} +\big)^2 \Big\rangle = \\
&\qquad = 4\Big\langle \big(\sum_{i\in P}\sum_{j\in T} h_{ij}\big)^2 \Big\rangle - 4\Big\langle \sum_{i\in P}\sum_{j\in T} h_{ij} \Big\rangle^2 \\
&\qquad = 4\sum_{i\in P}\Big\langle \big(\sum_{j\in T} h_{ij}\big)^2 \Big\rangle - 4\sum_{i\in P}\Big\langle \sum_{j\in T} h_{ij} \Big\rangle^2 = \\
&\qquad = 4\sum_{i\in P}\sum_{j\in T}\langle h_{ij}^2\rangle - 4\sum_{i\in P}\sum_{j\in T}\langle h_{ij}\rangle^2 = \\
&\qquad = 4\cdot d\cdot (n/2 - d)\cdot \mathrm{var}\ H
\end{aligned} \tag{21}$$

The variance can be calculated by averaging over these expressions weighted with the probability $p(d)$ for choosing two partitions with a distance d at random

$$p(d) = \lim_{s\to\infty} \phi_{sd} = \binom{n/2}{d}^2 \Big/ \binom{n}{n/2} \tag{22}$$

One obtains

$$2\mathrm{var}\ f = \sum_{d=0}^{n/2} d(n/2 - d)\binom{n/2}{d}^2 \cdot \mathrm{var}\ H \Big/ \binom{n}{n/2} \tag{23}$$

The d-dependent part of the sum can be calculated explicitly using $m = n/2$ and

$$\begin{aligned}
&\sum_{d=0}^{m} d\binom{m}{d}^2 = \frac{(2m-1)!}{[(m-1)!]^2}\frac{m}{2}\cdot\binom{2m}{m} \\
&\sum_{d=0}^{m} d^2\binom{m}{d}^2 = m^2\binom{2(m-1)}{m-1} = \frac{m^3}{2(2m-1)}\binom{2m}{m}
\end{aligned} \tag{24}$$

yielding finally

$$\mathrm{var}\ f = \frac{1}{2}\langle (f(x) - f(y))^2\rangle_{random} = \frac{n^2}{16}\cdot\frac{n-2}{n-1}\cdot \mathrm{var}\ H \tag{25}$$

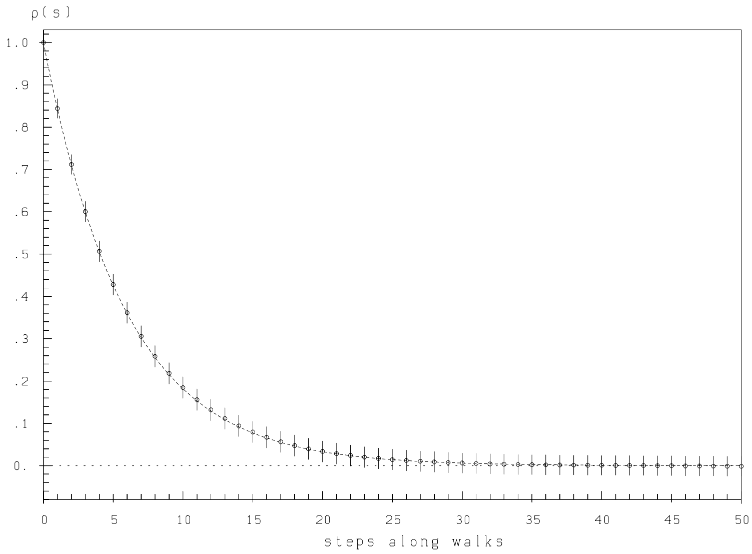

Figure 1: Typical example of an empirical autocorrelation function *r(s)* obtained by averaging 50 random walks of length 100000 on a graph with *n=50* vertices and connectivity *p=0.2*. The dotted line is the theoretical prediction equ.(29)

Thus the autocorrelation function is

$$\rho(d) = 1 - 8\frac{d}{n}(1+\frac{1}{n-2}) + 16(\frac{d}{n})^2(1+\frac{1}{n-2}) \tag{26}$$

independent of the distribution of the elements of the matrix H associated with the graph Γ.

Instead of treating $\rho(d)$ as a function of the distance of the landscape one may also use the autocorrelation function along a random walk as a characteristic of the landscape $\{x_s\}$

$$r(s) = \frac{\langle (f(x_t) - \langle f\rangle)\cdot(f(x_{t+s}) - \langle f\rangle)\rangle}{\text{var } f}. \tag{27}$$

Since the random walk is obviously ergodic and the distribution of values approaches a normal distribution by virtue of the central limit theorem, we expect that the "time series" sampled along the random walk is an AR(1) process (Weinberger 1990) in the limit of large n, and thus $r(s)$ at least approaches a single decaying exponential as $n \to \infty$. Surprisingly, there is no finite size correction to this behaviour: The two autocorrelation functions $\rho(d)$ and $r(s)$ are related by

$$r(s) = \sum_{d=0}^{s} \phi_{sd} \cdot \rho(d) \tag{28}$$

which, in the case of the GBP simplifies to

$$\begin{aligned} r(s) &= 1 - 8\frac{n-1}{n(n-2)} \cdot \Delta_1(s) + 16\frac{n-1}{n^2(n-2)} \cdot \Delta_2(s) = \\ &= (1 - \frac{8}{n} + \frac{8}{n^2})^s \\ &= \exp(-\frac{s}{\lambda^{walk}}) \end{aligned} \tag{29}$$

with the correlation length along the walk

$$\lambda^{walk} = \tau_2 = \frac{1}{8}(n-3) + \mathcal{O}(\frac{1}{n}) \tag{30}$$

Let us now define the scaled distance x by

$$x = d/\text{diam}\ \mathcal{C} = 2d/n \tag{31}$$

One obtains the scaled autocorrelation functions

$$\begin{aligned} \tilde{\rho}(x) &= 1 - 4x + 4x^2 - \tfrac{4}{n-2}(x - x^2) &&\to 1 - 4x + 4x^2 \\ \tilde{r}(x) &= (1 - \tfrac{8}{n}\tfrac{n-1}{n})^{\frac{8}{n} \cdot x/4} &&\to \exp(-\tfrac{x}{4}) \end{aligned} \tag{32}$$

which is exactly the same result as for the Sherrington Kirkpatrick spinglass in the limit $n \to \infty$ (Weinberger and Stadler 1992).

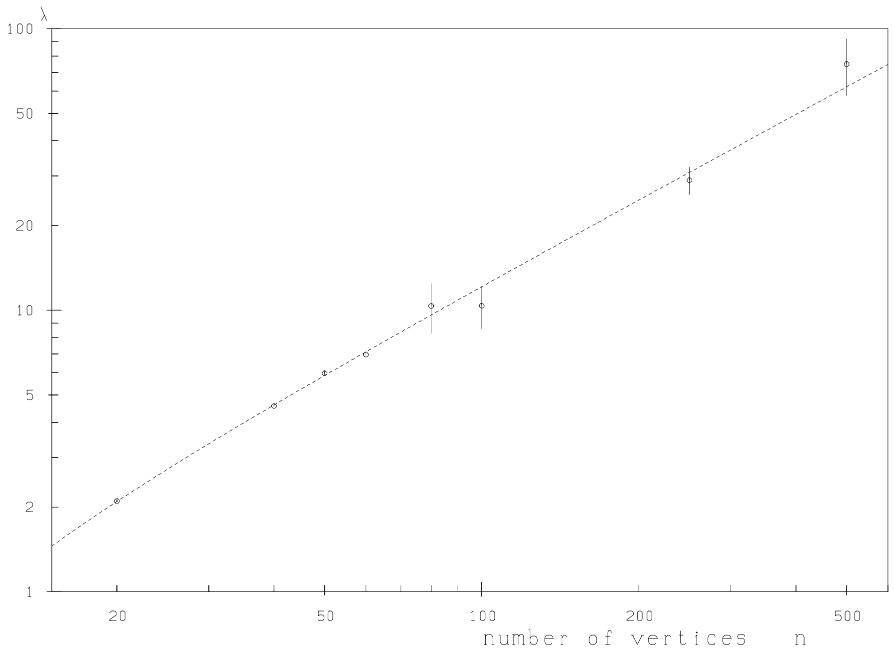

Figure 2: Empirical Correlation length λ^{walk} as function of the number of vertices n. The correlation length is, as predicted, independent of the connectivity p of the graph. The dotted line shows the exact prediction $\lambda^{walk}=-1/\ln(1-8/n+8/n^2)$.

4. Conclusions

The GBP may serve as a prototype model for fitness landscapes in combinatorial optimization and biological evolution (Fontana *et al.* 1991, 1992). It has a highly frustrated value landscape and is thus likely to be a typical problem. Nevertheless, some basic global properties — relaxation of random walks, pair correlation and expected value of the global optimum — can be obtained analytically.

It has been shown recently (Stadler and Schnabl 1992) for the Travelling Salesman Problem that the correlation structure of the landscape is closely related to the performance of optimization heuristics: The number of local optima decreases with increasing correlation length. In fact, there is a single local optimum in a patch with a radius of roughly the correlation length. Since trapping in local optima is the main reason for unsatiesfying performance of a heuristic, choosing a move set such as to maximize the correlation lenght and simultanously keeping constant the number of nearest neighbours seems to be a promising strategy for improving optimization algorithms.

It is interesting to note that, despite the fact that the configuration spaces $\mathcal{C}_{SK}$ Sherrington Kirkpatrick spin glass and $\mathcal{C}_{GBP}$ differ considerably, both systems exhibit the same correlation structure in the limit of large systems.

Acknowledgements

The authors wish to thank M. Eigen, J. McCaskill, K. Nieselt-Struwe, P. Schuster and E.D. Weinberger for stimulating discussions. Proofreading of the manuscript by John Yin and the generous supply of computer time by the *Computer Center of the University of Vienna*, supported by the IBM-EASI initiative, and the *Gesellschaft für Wissenschaftliche Datenverarbeitung, Göttingen* is gratefully acknowledged.

References

Banavar JR, Sherrington D and Sourlas N 1987 *J. Phys. A: Math. Gen.* **20** L1

Fontana W, Griesmacher T, Schnabl W, Stadler PF and Schuster P 1991 *Mh. Chem.* **122** 795

Fontana W, Stadler PF, Bauer E, Griesmacher T, Hofacker IL, Tacker M, Tarazona P, Weinberger ED and Schuster P 1992 subm. to *Phys.Rev.A*

Fu Y and Anderson PW 1986 *J. Phys. A: Math. Gen.* **19** 1605

Fu Y 1989 in: *Lectures in the Science of Complexity, SFI Studies in Science of Complexity*, ed. D Stein. Addison-Wesley, Longman.

Liao W 1987 *Phys. Rev. Lett.* **59** 1625

Sherrington D and Kirkpatrick S 1975 *Phys. Rev. Lett.* **35** 1792

Sorkin G 1988 *Combinatorial Optimization, Simulated Annealing and Fractals* IBM research report RC13674 (No. 61253)

Stadler PF and Schnabl W 1992 *Phys. Lett. A* **161** 337

Weinberger ED 1990 *Biol. Cybern.* **63** 325

Weinberger ED and Stadler PF 1992 *Why Some Fitness Landscapes are Fractal* (preprint)

Wiethege W and Sherrington D 1987 *J. Phys. A: Math. Gen.* **20** L9