1

Coloring in Sublinear Time

Andreas Nolte and Rainer Schrader
University of Cologne, Weyertal 80, 50931 Cologne, Germany

April 1999

We will present an algorithm, based on SA-techniques and a sampling procedure, that colors a given random 3-colorable graph with high probability in sublinear time. This result is the first theoretical justification of many excellent experimental performance results of Simulated Annealing [10, 17] applied to graph coloring problems.

1 Introduction

Johnson et al. [10] and Petford and Welsh [17] report on very good practical results of Simulated Annealing for coloring random graphs. Compared to deterministic algorithms such as described in [12] and [20], Simulated Annealing finds correct colorings faster on almost all random instances. Taking these practical results as a starting point, we will consider the same random model as in [17]. We assume a given set of n vertices partitioned into three color classes, each pair of vertices of different color being connected with a certain probability p. Thus, the random graphs are 3-colorable. Due to the simplicity of the representation, and due to the fact that in a sense the 3-colorability case is the most difficult, we consider only this case. However, most results seem to be extendible to the arbitrary k-colorability case.

There exist already some deterministic algorithms [3, 20, 12, 5] presented by various authors which color a given random instance correctly with three colors and with high probability. As usual, the term "with high probability" means that the probability tends to 1 as the problem size tends to infinity. All deterministic algorithms mentioned so far show the characteristic that

*Supported by DFG-grant (Schr 390/2-1)
Extended Version of [16].

the correct construction of the 3-coloring requires nearly equal sized color classes (up to a factor of $1+o(1)$). Moreover, this construction takes a number of steps, that is linear in the number of edges of a given random graph. In the usual sense the time complexity is of course optimal, because the size of the input is linear in the number of edges. Furthermore, a verification of the correctness of the coloring requires a number of steps that is at least as large as the number of edges.

But the extremely good performance of Simulated Annealing on random instances raises the question, wether it is possible to construct a correct 3-coloring of a given instance even faster. This means that we try to find a correct coloring without looking at every edge. Aside from the theoretical point of view, the answer of this question is of course only of practical value for on-line applications, where the demand of a correct coloring must only be fulfilled with high probability.

In the following section we will answer this question in the affirmative. We will describe an algorithm that uses SA-type techniques and a sampling procedure to compute the cost function more efficiently. This algorithm will stop after a number of steps that is strictly less than the number of edges. We will show that this algorithm produces with certainty a coloring of a given random graph with equal sized color classes and that this coloring is a correct one with high probability over all graphs and random steps.

Only two results are published by now concerned with the convergence of Simulated Annealing in polynomial time. Jerrum and Sorkin [9] proved the convergence to an optimal solution on certain random instances of the Graph Bisection Problem in $O(n^3)$ steps, where n is the number of vertices. Hajek and Sasaki [19] were engaged with the performance of Simulated Annealing applied to the Maximum Matching Problem and showed Simulated Annealing to be in a near optimal state in polynomial time. But these authors consider only the Metropolis process (Simulated Annealing at a certain fixed temperature) and thus investigate the convergence of a homogeneous Markov chain. In the following, we will show the convergence of Simulated Annealing with varying, time-dependent temperatures. Colorings in sublinear time are also found by the algorithm introduced by Prömel and Steger [18], however, they do not consider SA-type algorithms.

In order to improve the readability most proofs in this section are deferred to Chapter 9.

2 Random Model

We consider a quite simple random model ensuring that every graph taken from the induced probability space is 3-colorable with equal sized color

classes. The model introduced here is commonly accepted and already used by several authors [20, 17, 3] to analyze the performance of various coloring algorithms.

We consider a given set V of n vertices and color $n/3$ of these vertices red, $n/3$ of these vertices blue and $n/3$ of these vertices green (we assume n to be a multiple of 3 to get equal sized color classes). Then we connect vertices of different color with constant probability p getting an edge set $E \subset V^2$ (an edge is considered to be an unordered pair of vertices). In order to simplify the representation of our analysis we assume $p = 1/2$. However, the presented analysis in this chapter may be extended to some lower problem-dependent probability $p \leq 1/2$, but this is not considered further in the following.

After constructing the graph we forget the coloring. We call the induced probability space $\mathcal{G}$. The task is to color a given graph in $\mathcal{G}$ properly with 3 colors and high probability.

3 Idea of the Algorithm

The strategy to color a given $G \in \mathcal{G}$ is to recover the coloring that has been used to construct G. We take an arbitrary set A of $n/3$ vertices from the set of all vertices. Then we try to increase the imbalance in A by local exchanges. This implies that we try to increase the cardinality of the largest color class in A. After a certain small number of SA-transition steps we can guarantee a quite large imbalance. Then we use some hillclimbing steps (i.e. Simulated Annealing at temperature 0). In order to compute the cost function (i.e. the number of neighbors in A and $V \setminus A$) more efficiently we look only at a certain smaller sample of vertices in the corresponding sets. But, these samples are large enough to reflect the imbalance situation of A and $V \setminus A$. After these hillclimbing steps we can be sure that there are only vertices of two colors in A. Applying again SA-transitions followed by hillclimbing steps we will see that A will only contain vertices of one color. In order to 2-color $V \setminus A$ we apply a similar method as described above.

The main purpose of the following sections is to give a detailed description of the algorithm and a rigorous proof of the following main theorem in this chapter.

Theorem 1 *There exists a randomized algorithm using only SA-type transitions that determines a correct 3-coloring of a given $G \in \mathcal{G}$ in $O(n^\delta)$ steps ($\delta < 2$) with high probability over all graphs and random steps, while checking the correctness of the constructed coloring requires $\Omega(n^2)$ steps.*

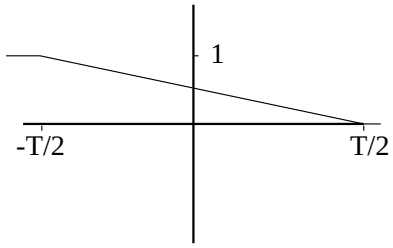

Figure 1: The acceptance function $a(x)$

4 Simulated Annealing

4.1 Local Exchanges

Let a graph $G \in \mathcal{G}$ be given. We partition the set of n vertices into a set A containing $n/3$ vertices and a set containing $2/3n$ vertices. For technical reasons we consider these partition as a permutation π with the first $n/3$ vertices belonging to $A = A_\pi$. The set of configurations of the following Simulated Annealing algorithm will therefore be the set of permutations π. We choose an arbitrary state π_0 as a starting state. To explain the transition steps let a state π_t in step t be given.

We choose a vertex u uniformly at random from the set A_{π_t} and a vertex v uniformly at random from the set $V \setminus A_{\pi_t}$. The proposed move is the change of the positions of u and v in π_t.

The cost function Δc of a proposed move considered here is the number of neighbors of v in A_{π_t} minus the number of neighbors of u in A_{π_t}. Thus, Δc measures the change of the number of edges between the vertices in A_{π_t}. The acceptance probability of a proposed move is

$$a(\Delta c) \quad = \quad \begin{cases} \frac{1}{2} - \frac{\Delta c}{T_t} & \text{for} \quad \Delta c \in [-T_t/2, T_t/2], \\ 0 & \text{for} \quad \Delta c > T_t/2, \\ 1 & \text{for} \quad \Delta c < -T_t/2, \end{cases}$$

where $T_t \in \mathbb{Q}$ is the temperature in step t. (We use rational temperatures due to complexity reasons.) The graph is outlined in Figure 1. The sequence $(T_t)_{t\in\mathbb{N}}$ is the sequence of decreasing real numbers known as cooling schedule of SA and will be specified in the following. This implies the following: If a proposed move decreases the number of edges between vertices in A_{π_t}, we accept it with a certain larger probability than a proposed move that increases the number of edges. It is intuitively easy to see that the current state π_t will be a state with a quite low number of edges in A_{π_t} provided that the number of steps is high enough. This acceptance function is not as common as $\min\{\exp\left(\frac{-\Delta c}{T_t}\right), 1\}$, but due to its symmetry it simplifies the following analysis.

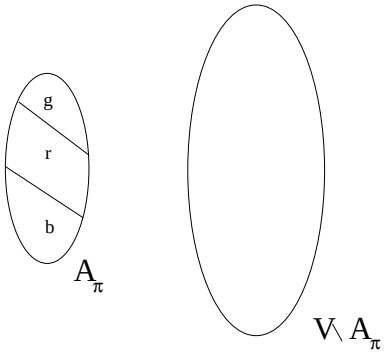

Figure 2: A partiton π

4.2 Expected Transition Probabilities

The key value that we use to analyze the performance of our local exchanges is the imbalance of a partition π. To explain this value let an arbitrary partition π be given. A_π is a set of $n/3$ vertices that are red, blue or green (known from the construction, but invisible for the algorithm). Let $r = n/9 + i_r$ be the number of red vertices, $b = n/9 + i_b$ be the number of blue vertices and $g = n/9 + i_g$ be the number of green vertices in A_π. Thus, we get $i_r + i_b + i_g = 0$. The imbalance i_π of the permutation π is the value of that i_* which corresponds to the largest color class in A_π. Therefore we get $i_\pi = \max\{i_r, i_b, i_g\}$ and $0 \leq i_\pi \leq 2/9n$ (see Figure 2).

The reason to consider the imbalance is that we have projected the quite complex coloring Markov chain on the natural numbers, a method analogous to that suggested by JERRUM and SORKIN [9]. This process is not necessarily Markovian any more, and this complicates the analysis, but the process can be bounded by dominating Markov chains on the natural numbers which are easy to analyze. First of all we try to get bounds for the expected value of i_π after one transition.

Proposition 1 *Let a partition π of a graph $G \in \mathcal{G}$ and a temperature $T_t > n^{1/2}$ be given. Let $u =$ "up" be the random variable on the set of all graphs that denotes the probability over all possible transitions to increase i_π in transition step t. $d =$ "down" denotes the corresponding random variable for a decrease of i_π. Then we get*
$\exists c_1, c_2 \in \mathbb{R}_+, k \geq 1/9$

$$E(u) \geq k - c_1 \frac{i_\pi}{n}$$

and

$$E(d) \leq k + c_1 \frac{i_\pi}{n} - c_2 \min\left\{\frac{i_\pi}{T_t}, 1\right\},$$

where E denotes the expectation over all graphs $G \in \mathcal{G}$.

Proof: See Chapter 9. □

Assuming $T_t \ll n$, this proposition states that the imbalance of the current partition is more likely to be increased than to be decreased, and this depends crucially on the size of the current imbalance. This is a positive hint that it is possible to reach our target getting a partition with a quite high value of the imbalance after a certain number of steps.

But so far, we know only some facts about the expected values of the transition probabilities. In the following section we are interested in the deviation of the current transition probabilities from the expected value.

4.3 Deviation

To measure the deviation from the expected values of u and d we need the following definition

Definition 2 *A pair (G, π) of a graph G and a partition π is called* α-deviant*, if either $u = u(G, \pi)$ or $d = d(G, \pi)$ differs from its expected value (over all graphs) by more than α.*

The following lemma is the first result we can get about α-deviation by a direct application of the method of bounded differences.

Lemma 3 *Let a partition π and an $\alpha > 0$ be given. The probability over all graphs $G \in \mathcal{G}$ that a pair (G, π) is α-deviant is at most $2\exp(O(\alpha^2 {T_t}^2))$.*

Proof: See Chapter 9. □

But it turns out that this lemma alone is, although necessary in the following, not sufficient to get proper bounds of the deviation. These proper bounds are necessary to ensure that the current transition probabilities do not differ too much from their expected values during a certain number of subsequent steps. In the following we apply an idea of Jerrum and Sorkin [9] to improve the bound of the lemma.

The key idea is to introduce a new quantity that is a suitable sum of transition probabilities. We will see that we can get bounds for this aggregated value allowing us to derive better bounds for the whole process than

looking at the individual transition probabilites of every partition π as in Lemma 3. For a given graph G let $P(\pi, \pi')$ be the generation and acceptance probability for the transition from π to π', and $EP(\pi, \pi')$ be the expected value over all graphs. The crucial value considered here is the sum over all partitions π' of the difference $|P(\pi, \pi') - EP(\pi, \pi')|$.

First of all we try to get an analogue of Lemma 3 for the new quantitiy.

Proposition 4 *Let a temperature $T_t > 0$ be given. With high probability over all graphs and for all partitions π we get*

$$\sum_{\pi'} |P(\pi, \pi') - EP(\pi, \pi')| = O\left(\frac{\sqrt{n}}{T_t}\right).$$

Proof: See Chapter 9. □

With the help of the last proposition we can prove the main theorem of this section that bounds the deviation of u and d. The proof idea goes back to JERRUM and SORKIN [9], who solved randomized instances of the Graph Bisection Problem via the Metropolis process.

Theorem 2 *For any temperature $0 < T < poly(n)$ and any $\alpha > 0$, the Simulated Annealing process at constant temperature T for $t = \Theta(\alpha^2 T^3/\sqrt{n}\log^2 n)$ steps encounters a α-deviant state with probability of at most*

$$\exp(-\Omega(\alpha^2 T^2/\log^2 n)).$$

Proof: See Chapter 9. □

4.4 Random Walks

In this section we analyze the time dependent behavior of the imbalance of the current partition during the Simulated Annealing process. We can portray this value as a projection of a Markov process on the natural numbers. This process depends on a hidden variable, namely the current partition. Thus, it is not necessarily Markovian. But, with the help of the results concerning the expected transition probability and the deviation we can construct a process that is a lower bound of the process of the imbalance values. This means that we can construct a new random walk with constant transition probabilities being quite easy to analyze. Additionally, we can be

sure that our process of interest ($(i_{\pi_t})_{t\in\mathbb{N}}$) has got "on the average" larger values than our random walk. The technique used to carry out this comparison is known as coupling and is made precise in the next proposition. A similar version can be found in [9].

Proposition 5 *Let $(X_t)_{t\in\mathbb{N}}$ be a Markov chain with state space Ω and $\Xi : \Omega \to \mathbb{N}$ be a projection on the natural numbers. Suppose $(\xi_t)_{t\in\mathbb{N}}$ to be a random walk on the natural numbers having $\xi_{t+1} \in \{\xi_t - 1, \xi_t, \xi_t + 1\}$ as its only allowed transitions. Suppose further that it is a probabilistic lower bound of the process $\Xi(X_t)_{t\in\mathbb{N}}$ in the following sense*

$$\begin{aligned}
&\xi_0 \leq \Xi(X_0)\\
&P(\xi_{t+1} = s+1 | \xi_t = s) \leq P(\Xi(X_{t+1}) \geq s+1 | \Xi(X_t) = s)\\
&P(\xi_{t+1} = s-1 | \xi_t = s) \geq P(\Xi(X_{t+1}) = s-1 | \Xi(X_t) = s)\\
&P(\Xi(X_{t+1}) < s-1 | \Xi(X_t) = s) = 0
\end{aligned}$$

for $s \in \mathbb{N}$ arbitrary.

Then there exist a coupling $(Y_t)_{t\in\mathbb{N}}$ of $(\xi_t)_{t\in\mathbb{N}}$ and a coupling $(Z_t)_{t\in\mathbb{N}}$ of $(\Xi(X_t))_{t\in\mathbb{N}}$ (i.e. Y_t (Z_t) has the same distribution as ξ_t ($\Xi(X_t)$)) with $Y_t \leq Z_t$ for all $t \in \mathbb{N}$.

Proof: See Chapter 9. □

Now, we are able to analyze the stochastic process i_{π_t}. Let $0 < \gamma \leq 1/100$ be fixed. We choose a temperature $T_0 = \lceil n^{2/3-\gamma} \rceil$ and keep it fixed during the first r steps (r will be defined in the following). By defining $\alpha = n^{-1/3}$ we can conclude with the help of Proposition 2 that the imbalance process i_{π_t} does not reach a partition with a larger deviation than α with high probability during the first $\lceil n^{5/6-4\gamma} \rceil$ steps. In the following we concentrate on this case. We define a random walk $(lb_t^b)_{t\in\mathbb{N}}$ on the natural numbers, starting at 0, that plays the role of the lower bound of i_{π_t} at the beginning in the interval $[0, n^{1/3}]$

$$\begin{aligned}
&lb_0^b = 0,\\
&lb_{t+1}^b \in \{lb_t^b, lb_t^b + 1, lb_t^b - 1\},\\
&P(lb_{t+1}^b = n^{1/3} + 1 | lb_t^b = n^{1/3}) = P(lb_{t+1}^b = -1 | lb_t^b = 0) = 0
\end{aligned}$$

and

$$\begin{aligned}
&P(lb_{t+1}^b = s+1 | lb_t^b = s) = k - 2\alpha \qquad \text{for} \quad s \in \{0, \ldots, n^{1/3} - 1\},\\
&P(lb_{t+1}^b = s-1 | lb_t^b = s) = k + 2\alpha \qquad \text{for} \quad s \in \{1, \ldots, n^{1/3}\}
\end{aligned}$$

with k chosen as in Proposition 1. Using Proposition 1 and the facts that $i_\pi \geq 0$ and only a deviation of α is allowed, we can see that lb_t^b and the process i_{π_t} are correlated in the same way as ξ_t and $\Xi(X_t)$ in Proposition 5. Thus, lb_t^b is in fact a probabilistic lower bound of i_{π_t} in the sense of Proposition 5.

Now, we analyze the time being necessary for the process lb_t^b to reach $n^{1/3}$. It is a random walk with one elastic barrier 0 and can be analyzed quite easily with standard Markov chain theory.

Lemma 6 *Let $\epsilon > 0$ be given. The random walk lb_t^b will hit $n^{1/3}$ with high probability within $O(n^{2/3+\epsilon})$ steps.*

Proof: See Chapter 9. □

According to Proposition 5 we obtain - with the help of the above described identification of lb_t^b with ξ_t and i_{π_t} with $\Xi(X_t)$ - that there exist couplings Y_t and Z_t of lb_t^b and i_{π_t} with $Y_t \leq Z_t$ for all $t \in \mathbb{N}$. Since Y_t has the same distribution as lb_t^b, we can derive the same result as in Lemma 6 for Y_t. Therefore, we can be sure that Z_t has hit a state $s \geq n^{1/3}$ with high probability within $O(n^{2/3+\epsilon})$ steps. Due to the fact that Z_t and i_{π_t} have got the same distribution, we get

Lemma 7 *The process i_{π_t} will hit a state $s \geq n^{1/3}$ with high probability within $O(n^{2/3+\epsilon})$ steps.*

In the following we consider the behavior of the process i_{π_t} on the interval $[n^{1/3}, n^{1/2+\gamma}]$. Again, we construct a random walk on the natural numbers playing the role of a lower bound for i_{π_t}. Strictly speaking, it is a series of random walks lb_t^j, $j \in \{0, \ldots, \lfloor \frac{1}{6\gamma} \rfloor\}$, each being a lower bound for i_{π_t} on $[\frac{n^{1/3+j\gamma}}{2}, n^{1/3+(j+1)\gamma}]$. Let

$lb_0^j = n^{1/3+j\gamma}$,
$lb_{t+1}^j \in \{lb_t^j, lb_t^j + 1, lb_t^j - 1\}$,

$$\begin{aligned} & P(lb_{t+1}^j = n^{1/3+(j+1)\gamma} + 1 | lb_t^j = n^{1/3+(j+1)\gamma}) \\ = \;& P(lb_{t+1}^j = n^{1/3+j\gamma}/2 + 1 | lb_t^j = n^{1/3+j\gamma}/2) \\ = \;& 0 \end{aligned}$$

and

$P(lb_{t+1}^j = s+1 | lb_t^j = s) = k - \Theta(n^{-1/3})$ for $s \in \{n^{1/3+j\gamma}/2, \ldots, n^{1/3+(j+1)\gamma} - 1\}$,
$P(lb_{t+1}^j = s-1 | lb_t^j = s) = k - \Theta(n^{-1/3+(j+1)\gamma})$ for $s \in \{n^{1/3+j\gamma}/2 + 1, \ldots, n^{1/3+(j+1)\gamma}\}$.

Due to Proposition 1 and the allowed deviation of $\alpha = n^{-1/3}$ the processes lb_t^j and i_{π_t} are also in the same correlation as ξ_t and $\Xi(X_t)$. Using Proposition 5 it is again sufficient to analyze lb_t^j.

Lemma 8 *Let $j \in \{0, \ldots, \lfloor \frac{1}{6\gamma} \rfloor\}, \epsilon > 0$ be given. The process lb_t^j starting at $n^{1/3+j\gamma}$ will hit $n^{1/3+(j+1)\gamma}$ earlier than $\frac{n^{1/3+j\gamma}}{2}$ with high probability and within $O(n^{2/3+\epsilon})$ steps.*

Proof: See Chapter 9. □

Since γ was fixed, and lb_t^j and i_{π_t} are correlated in a way allowing us to apply Proposition 5, we get with the same coupling argument described above the following corollary.

Corollary 9 *Let $\epsilon > 0$ be given. The process i_{π_t} will hit $n^{1/2+\gamma}$ with high probability within $O(n^{2/3+\epsilon})$ steps.*

After $ln^{2/3+\epsilon}$ steps with a suitable constant $l \in \mathbb{N}$ we stop the process (a rigorous computation with exact factors instead of the O/Ω-notation yields, that $l = 9$ is sufficient). Carrying out a similar analysis to that above with a random walk on $[\frac{n^{1/2+\gamma}}{2}, n^{3/4}]$ that starts at $n^{1/2+\gamma}$, we get a random walk on the natural numbers that is a lower bound of i_{π_t}. It reaches $n^{3/4}$ earlier than $\frac{n^{1/2+\gamma}}{2}$ with high probability. Because reaching $n^{3/4}$ would take at least $n^{3/4}$ steps, we could derive that after stopping, our process i_{π_t} is in a state somewhere between $\frac{n^{1/2+\gamma}}{2}$ and $n^{3/4}$.

Now we lower the temperature T_t. The new temperature will be $T_t = \lceil n^{1/2+\gamma} \rceil$ for $t \in \{9\lceil n^{2/3+\epsilon} \rceil, \ldots\}$ and $\gamma > 0$ fixed. Setting $\alpha = n^{-1/8}$ we can derive with Proposition 2 that no α-deviant state will occur in the next $n^{3/4+3\gamma}$ steps with high probability.

To analyze the imbalance process in the interval $[n^{1/2+\gamma}/4, n^{3/4}]$ we define the lower bound process lb_t^e

$lb_0^e = n^{1/2+\gamma}/2$,
$lb_{t+1}^e \in \{lb_t^e, lb_t^e + 1, lb_t^e - 1\}$,
$P(lb_{t+1}^e = n^{3/4}+1 | lb_t^e = n^{3/4}) = P(lb_{t+1}^e = n^{1/2+\gamma}/4 - 1 | lb_t^e = n^{1/2+\gamma}/4) = 0$

and

$$P(lb_{t+1}^e = s+1 | lb_t^e = s) = k_1 \quad \text{for} \quad s \in \{n^{1/2+\gamma}/4, \ldots, n^{3/4}-1\},$$
$$P(lb_{t+1}^e = s-1 | lb_t^e = s) = k_2 \quad \text{for} \quad s \in \{n^{1/2+\gamma}/4+1, \ldots, n^{3/4}\}$$

with constants $k_1 < k_2$ suitably chosen according to Proposition 1.

Carrying out an analysis in the same way as in the proof of Corollary 9 and proving as described after Corollary 9 that the process will stay above $n^{3/4}$ we get the main result of this section.

Proposition 10 *Let $\gamma, \epsilon > 0$ be sufficiently small constants. Running the Simulated Annealing process for $9n^{2/3+\epsilon}$ steps with temperature $T = n^{2/3-\gamma}$ and $16n^{3/4+\epsilon}$ steps with temperature $T = n^{1/2+\gamma}$ yields a partition π with an imbalance value i_π of at least $n^{3/4}$.*

5 Hillclimbing

Assuming a given graph G and a partition π with imbalance $i_\pi \geq n^{3/4}$, we will define in this chapter certain hillclimbing steps (i.e. Simulated Annealing at temperature 0) in order to remove one color class in A_π. Roughly speaking we will compare the number of neighbors of a randomly chosen vertex in A_π and $V \setminus A_\pi$. To save time during the computation of the cost function of one move (i.e. the number of neighbors) it is not necessary to look at the complete sets. We choose only a sample in A_π and $V \setminus A_\pi$ of size $n^{3/4-\epsilon}$ ($\epsilon > 0$ sufficiently small) uniformly at random.

In the following we will see that the imbalance situation will be reflected in the samples. Intuitively it is clear that a vertex having the same color as the minimum color class in A_π will tend to have more neighbors in the sample in A_π than in the sample in $V \setminus A_\pi$.

This idea is made precise in the following. First we consider the case of choosing a sample of size $n^{3/4-\epsilon}$ uniformly at random from A_π.

Lemma 11 *Let $0 < \epsilon < 1/24$ and a partition π with imbalance $i_\pi \geq n^{3/4}$ be given. The number of vertices of a distinct color $*$ in a sample of $n^{3/4-\epsilon}$ vertices chosen uniformly at random in A_π differs only within a range of $\Delta = O(n^{1/2-2\epsilon})$ from the expected number*

$$n^{3/4-\epsilon}\left(\frac{n/9 + i_*}{n/3}\right)$$

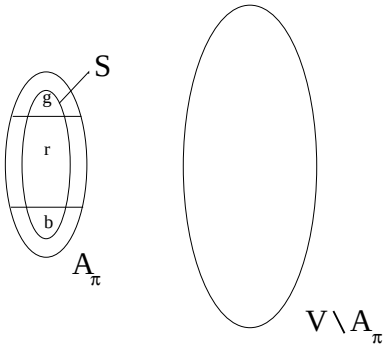

Figure 3: The sample S

of the corresponding independent Bernoulli trials with high probability over all chosen samples.

Proof: See Chapter 9. □

Using Lemma 11 we can derive that after the local exchanges in the last chapter our sampling process yields a reliable smaller copy of our partition with high probability.

Again, let a graph G and a partition with imbalance $n^{3/4}$ be given. Furthermore, we assume a given reliable sample S with error bound Δ as in Lemma 11. Thus, we have the situation as in Figure 3.

Let v be a vertex of the largest color class in our sample, which we assume to be red. The expected number of neighbors over all graphs in S is

$$\begin{aligned} 1/2(S_g + S_b) &\leq 1/2n^{3/4-\epsilon}\left(\frac{n/9 + i_g + n/9 + i_b}{n/3}\right) + \Delta \\ &\leq 1/3n^{3/4-\epsilon} - 3/2n^{1/2-\epsilon} + \Delta \\ &\leq 1/3n^{3/4-\epsilon} - 4/3n^{1/2-\epsilon} \end{aligned}$$

with S_b and S_g being the number of green and blue vertices in S. Assuming green to be the least frequent color in the partition we get the following bound for the expected number of neighbors of a green vertex in S

$$\begin{aligned} 1/2(S_b + S_r) &\geq 1/2n^{3/4-\epsilon}\left(\frac{n/9 + i_b + n/9 + i_r}{n/3}\right) - \Delta \\ &\geq 1/3n^{3/4-\epsilon} + 3/2n^{-1/4-\epsilon}(i_b + i_r) - \Delta \\ &\geq 1/3n^{3/4-\epsilon} + 3/4n^{1/2-\epsilon} - \Delta \\ &\geq 1/3n^{3/4-\epsilon} + 1/2n^{1/2-\epsilon} \end{aligned}$$

with S_r being the number of red vertices in S.

Now we consider the case of choosing a sample of size $n^{3/4-\epsilon}$ uniformly at random in $V \setminus A_\pi$. One can show analogously:

The number of neighbors of a red vertex in this sample is at least

$$1/3n^{3/4-\epsilon} + 1/2n^{1/2-\epsilon}.$$

with high probability. As an upper bound for the number of neighbors of a green vertex we get:

$$1/3n^{3/4-\epsilon} - 1/2n^{1/2-\epsilon}.$$

In the following we will describe the hillclimbing steps:

1. Take one of the sets $\{A_\pi, V \setminus A_\pi\}$ as starting set B
2. Choose a vertex v uniformly at random in B
3. Choose a sample of size $n^{3/4-\epsilon}$ uniformly at random in A_π and $V \setminus A_\pi$
4. Compare the number of neighbors in both samples
5. If the number of neighbors in the sample in B is greater than the number of neighbors in the other sample, we move v from B to the complement, take B as this complement and go to 2.
   else go to 2.

Obviously we can be sure that our sizes of the sets remain constant (up to +/-1).

With the analysis at the beginning of this chapter we can see that a red vertex will readily move to or stay at the smaller set. In contrast to that, a green vertex will readily move to or stay at the bigger set. Moreover it is easy to see that the bias towards these directions will be reinforced during the process so that the above described behavior will continue during the hillclimbing process. After $O(n \log n)$ steps every vertex is visited at least once with high probability. This implies that there are only vertices of two colors left in the smaller set.

6 Simulated Annealing again

After the hillclimbing steps we start our initial SA-algorithm (see Chapter 4) again. To analyze this process we will introduce a new value i_π^2 similar to our initial imbalance, but this time suitable for the two color case. Let red and blue be the two colors in A_π and $r = n/6 + i_r$ and $b = n/6 + i_b$ be the corresponding cardinalities. Thus we get $i_r + i_b = 0$. The imbalance i_π^2 is the value of that i_* which corresponds to the largest color class in A_π. Therefore we get $i_\pi^2 = \max\{i_r, i_b\}$ and $0 \leq i_\pi \leq 1/6n$. We are again interested in the number of steps necessary to ensure that $i_\pi^2 \geq n^{3/4}$ with high probability over all graphs and transition steps.

We consider the case $i_\pi^2 < n^{7/8}$ first. Let green denote the color of those vertices that are not n the smaller set. With the help of Chernoff bounds it can be easily seen that the number of neighbors in the smaller set A_π is $\Omega(n)$ with high probability. Because the absolute value of the cost difference Δc will be greater than $T/2$ for our choices of temperatures $T_t \in \{n^{2/3-\gamma}, n^{1/2+\gamma}\}$, this implies the following:

A transition with a chosen green vertex in $V \setminus A_\pi$ will not be accepted. Therefore we can be sure that in this case no green vertex will move to A_π.

Choosing k random pairs of vertices (u, v) with $u \in A_\pi$ and $v \in V \setminus A_\pi$ we have to ensure that among these pairs there are enough pairs with no green vertices involved. The reason for this is of course that these transitions will not help us to improve the imbalance in A_π. But the number of pairs (u, v) with a green vertex v is Bernoulli distributed with probability 1/2. Using Chernoff bounds one can easily see that there will be at least $k/4$ pairs with no green vertices involved.

Using now temperature $T = n^{2/3-\gamma}$ for the first $4n^{2/3-\gamma}$ steps and $T = n^{1/2+\gamma}$ for the next $72n^{3/4+\epsilon}$ steps one can prove with an analogous analysis to that in Chapter 4 that the process will reach an imbalance of at least $n^{3/4}$. Taking care of the case of an initial imbalance of $i_\pi^2 \geq n^{7/8}$ we can summarize as a result of our second SA-application:

Proposition 12 *Let ϵ, γ be sufficiently small constants. Running the SA-process for $4n^{2/3-\epsilon}$ with temperature $T = n^{2/3-\gamma}$ and $72n^{3/4+\epsilon}$ steps with temperature $T = n^{1/2+\gamma}$ yields a partition π with*

- $|\{\textit{red vertices in } A_\pi\}| \geq n/6 + n^{3/4}$
- $|\{\textit{blue vertices in } A_\pi\}| \leq n/6 - n^{3/4}$
- $|\{\textit{red vertices in } A_\pi\}| \leq 100n^{3/4}$

With the help of the last proposition we can be sure that after our second application of the SA algorithm the largest color class in A_π dominates the other color classes heavily. Applying the hillclimbing algorithm ($O(n\log(n)$ steps) again yields a set A_π, that contains only vertices of one color class.

6.1 2-coloring

After separating one color as described in the last section we have to 2-color the remaining $2/3n$ vertices. This is not very difficult, since it is very easy to find an $O(n^2)$ algorithm that colors the remaining vertices correctly. But in order to get an overall performance that is strictly less than the quadratic bound we have to be more careful. The obvious idea to get better performance bounds is to apply the same idea as used in the last section to separate the vertices of one color from the rest.

Starting with an arbitrary bisection of the remaining vertices, we try to increase again the imbalance of an arbitrary chosen bisection with SA steps. After carrying out a number of local exchanges we can separate the two remaining color classes by applying the same hillclimbing steps as described above.

Input: arbitrary chosen bisection $\pi = (\pi_1, \pi_2)$ of the remaining $2/3n$ vertices, ϵ, δ, γ sufficiently small.

Output: 2-coloring of the remaining $2/3n$ vertices.

1. Execute $9n^{2/3+\epsilon}$ steps of the SA local exchanges with temperature $T = n^{2/3-\gamma}$.
2. Execute $16n^{3/4+\epsilon}$ steps of the SA local exchanges with temperature $T = n^{1/2+\gamma}$ and obtain a partition of the vertices in two equally sized sets V_1 and V_2.
3. Apply $2n\log(n)$ steps of the hillclimbing algorithm to the two sets of vertices V_1, V_2.

Because the analysis of this part of the coloring is analogous to the analysis carried out in the last sections we omit the detailed proofs here.

7 Concluding remarks

We have proved that our algorithm using only SA-type techniques will converge to a proper coloring in sublinear time with high probability. Although

the transition steps of our algorithm are quite simple one may wonder if they can be simplified further.

PETFORD and WELSH [17] suggested the following algorithm, taking the same random model as used above. Choose an arbitrary coloring as starting state. Then choose a vertex and a color uniformly at random. The cost function is the number of wrong edges, i.e. edges between vertices of the same color, that are forbidden in a proper coloring. They show by some experimental results that the algorithm works well in practice but a theoretical ananlysis is still missing.

8 Probability Theory

Because estimates of sums of random variables are quite important in this paper, we list a few in the following.

A very important special case a random variable f being Bernoulli distributed. This means $\Omega = \{0, 1\}$ and $P(f = 1) = p$ with a certain probability p and $P(f = 0) = 1 - p$. In the following we call a sum of Bernoulli random variables binomial ($B(n, p)$) distributed. One of the first published bounds is due to ANGLUIN and VALIANT [2]:

Proposition 13 (ANGLUIN *and* VALIANT) *If* $Y \in B(n, p)$, *then for all* $\alpha, 0 < \alpha < 1$

$$P(Y \leq (1 - \alpha)np) < e^{-\alpha^2 np/2} \quad \textit{and} \quad P(Y \geq (1 + \alpha)np) < e^{-\alpha^2 np/3}.$$

□

Sometimes a version of CHERNOFF is easier to apply

Proposition 14 (CHERNOFF) *Let* $f_1, \ldots, f_n$ *be a sequence of independent Bernoulli trials with* $P(f_i = 1) = p_i$ *and* $P(f_i = 0) = 1 - p_i$. *Define* $Y = \sum f_i$, *so that* $E(Y) = \sum p_i$. *Then for* $\beta \in [0, 1]$

$$P(|Y - E(Y)| > \beta E(Y)) \leq 2 \exp(-0.38\beta^2 E(Y)).$$

□

In addition to Proposition 14, a certain generalisation of the CHERNOFF bounds known as the method of bounded differences is required in our analysis in this thesis. Here the f_i need not to be Bernoulli distributed, but the range of these values must be bounded in a certain way.

Proposition 15 *Let $f_1, \ldots, f_n$ be independent random variables with f_k taking values in a set A_k for each k. Suppose that the (measurable) function $f : \prod A_k \to \mathbb{R}$ satisfies $|f(x) - f(x')| \leq c_k$, whenever the vectors x and x' differ only in the k-th coordinate. Let Y be the random variable $f(f_1, \ldots, f_n)$. Then for any $t > 0$*

$$P(|Y - E(Y)| \geq t) \leq 2\exp(-2t^2 / \sum c_k^2).$$

□

Before we come to the analysis of random walks we have to state an inequality concerning convex functions known as the inequality of JENSEN.

Proposition 16 *Let $J \subset \mathbb{R}$ be an interval, $g : J \to \mathbb{R}$ a convex function and $f : \Omega \to J$ a random variable. Then*

$$E(g(f)) \geq g(E(f))$$

Proof: See FELLER [6]. □

In the rest of this section we will be concerned with the analysis of random walks on the natural numbers. This represents another important proof technique used in this paper.

To give a formal framework we will define a random walk $(Y_t)_{t \in \mathbb{N}}$ on the integers by

$$\begin{gathered} Y_0 = z \\ P(Y_{t+1} = l + 1 | Y_t = l) = p \\ P(Y_{t+1} = l - 1 | Y_t = l) = q \\ P(Y_{t+1} = l | Y_t = l) = 1 - p - q \end{gathered}$$

with suitable constants $p, q \in [0, 1], p + q \leq 1$ and arbitrary $z \in \mathbb{Z}$.

A gambler could interprete this random walk as the amount of money that he owns after the t'th stake. He starts with an initial capital z, and in each step he can win a dollar with probability p and lose a dollar with probability q.

First of all we will consider the following special case with 0 being a reflecting barrier. That means

$P(Y_t < 0) = 0$
$P(Y_{t+1} = 1|Y_t = 0) = 1 - P(Y_{t+1} = 0|Y_t = 0).$

Thus, the random walk cannot become negative, if $z \geq 0$. We are interested in the expected time to hit a certain fixed number a for the first time. This value is the so called expected first hitting time D_z, starting at $z \leq a$. According to Feller [6] (pp. 344) the general solution for D_z is

$$D_z = \frac{z}{q-p} + A + B\left(\frac{q}{p}\right)^z$$

with A and B being constants (depending on a), that must fit the boundary conditions

$$D_a = 0 \quad \text{and} \quad D_0 = 1 + pD_1 + (1-p)D_0.$$

Solving for A and B and substituting yields

$$D_0 = -\frac{q}{q-p} + \frac{q}{(q-p)^2}\left[\left(\frac{q}{p}\right)^z - 1\right],$$

which is used frequently in the following chapter during the analysis of random walks with one reflecting barrier.

The next case that we want to consider is the following. There are no elastic barriers any more, and we start our random walk at a certain number z. Moreover, we are given two integral numbers a and b with $a \leq z \leq b$, and we want to answer the question which of the two numbers the random walk will touch first. Additionally, we want to calculate the expected time of this process. According to Feller [6] the probability that the event $Y_t = a$ occurs before $Y_t = b$ equals

$$\frac{\left(\frac{q}{p}\right)^{z-a} - \left(\frac{q}{p}\right)^{b-a}}{1 - \left(\frac{q}{p}\right)^{b-a}}.$$

The expected time to hit either a or b is

$$-\frac{z-a}{q-p} + \frac{b-a}{q-p}\left[\frac{1 - \left(\frac{q}{p}\right)^{z-a}}{1 - \left(\frac{q}{p}\right)^{b-a}}\right].$$

9 Proofs

Proof of Proposition 1:
Let $i_\pi = i_r$. This implies red to be the largest color class in our set A_π. First of all we want to determine the probability of a proposed move to increase or decrease i_r. Obviously, the probability to propose an i_r-increasing and i_b-decreasing move is

$$\frac{9b(n/3-r)}{2n^2}.$$

Symmetrically we get for the probability to decrease i_r and increase i_b

$$\frac{9r(n/3-b)}{2n^2},$$

and analogous values for the exchanges of red and green vertices.

Now we concentrate on the acceptance probabilities. Let an i_b-increasing and i_r-decreasing move be given. This is a bad move that could possibly decrease the imbalance i_π. C_d denotes the change of the cost function after applying this move to the partition π. Therefore, C_d is a random variable on the set of all graphs G.

Thus, C_d can be expressed as the sum and difference of $n/3+g$ independent Bernoulli random variables B_p, which correspond to the $n/3+g$ edges affected by the i_b-increasing and i_r-decreasing move.

$$\begin{aligned} C_d &= gB_p + rB_p - gB_p - bB_p \\ &= -gB_p + gB_p - (n/9 + i_b)B_p + (n/9 + i_r)B_p \end{aligned}$$

We have omitted the sum notation to avoid confusion about too many indices. kB_p denotes the sum of k independent Bernoulli trials with expectation p.

Now we define a new, symmetrically distributed random variable C_0 that we want to compare with C_d. We take the first $2g$ terms in C_d and define them as the first $2g$ terms in C_0 without changing any signs. Then we look at the following $2/9n + i_r + i_b$ terms of C_d. We define the first $\lfloor 1/2(2/9n + i_r + i_b)\rfloor$ of these terms as the next $\lfloor 1/2(2/9n + i_r + i_b)\rfloor$ terms with *negative* sign in C_0 and the rest of the terms as the last terms in C_0, but with *positive* sign. Thus we get

$$C_0 = -gB_p + gB_p - \lfloor 1/2(2/9n + i_r + i_b)\rfloor B_p + \lceil 1/2(2/9n + i_r + i_b)\rceil B_p$$

and it follows that

$$C_0 - C_d \leq (i_b - i_r)B_p \leq 0.$$

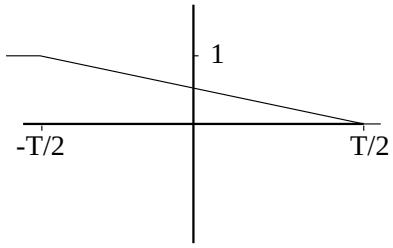

Figure 4: The acceptance function $a(x)$

A short calculation of the variance σ of C_0 yields

$$\sigma^2(C_0) = O(n) \tag{1}$$

To estimate the expected difference between $a(C_d)$ and $a(C_0)$ we need the following technical lemma.

Lemma 17 *Let* $M = \{G \in \mathcal{G} || C_0| \leq n^{1/2+\lambda}, C_d - C_0 \geq 1/20(i_r - i_b)\}$. *Then* $P(M) = \Omega(1)$ *for any* $\lambda > 0$ *follows.*

Proof: Using the Chebycheff inequality we get

$$P(|C_0| \leq n^{1/2+\lambda}) \geq 1 - \frac{1}{n^{1/2+\lambda}}\sigma(C_d),$$

which tends to 1 for $n \to \infty$. Furthermore, with the help of the inequality of Angluin and Valiant we obtain

$$P(C_d - C_0 \geq 1/20(i_r - i_b)) \geq 0,55$$

for $i_r - i_b \geq 4$. (For $i_r - i_b \in \{0, 1, 2, 3\}$ we get the claim by a direct calculation.) By combining the last two inequalities the Lemma 1 follows. □

Applying the acceptance function $a(x)$ (see Figure 4) we obtain

$$\begin{aligned} E(a(C_d)) - E(a(C_0)) &= E(a(C_d) - a(C_0)) \\ &= \int_M a(C_d) - a(C_0) + \int_{\bar{M}} a(C_d) - a(C_0) \\ &\leq \int_M a(C_d) - a(C_0), \end{aligned} \tag{2}$$

because $C_0 \leq C_d$ and $a(x)$ is monotonously decreasing.

Let a $G \in M$ be given. A standard analysis argument yields

$$\frac{a(C_d) - a(C_0)}{C_d - C_0} = a'(\delta)$$

with $\delta \in [C_0, C_d]$, if $C_d \leq T_t/2$. Moreover we obtain

$$a(C_d) - a(C_0) \leq -1/4 \tag{3}$$

for $C_d > T/2$ and $T_t \geq 4n^{1/2+\epsilon}$.

Therefore

$$a(C_d) - a(C_0) = a'(\delta)(C_d - C_0) = \frac{1}{T_t}(C_0 - C_d) \leq \frac{1}{20}\frac{i_b - i_r}{T_t} \tag{4}$$

with $C_d \leq T_t/2$. With Inequalities (2), (3) and (4) we obtain

$$E(a(C_d)) - E(a(C_0)) \leq -\frac{1}{20}\min\left\{\frac{i_r - i_b}{T_t}, 1\right\}. \tag{5}$$

This means that the acceptance probability of the bad move, which decreases the number of red vertices and increases the number of blue vertices, is by $\Omega(\min\{\frac{i_r - i_b}{T_t}, 1\})$ smaller than our reference value $E(a(C_0))$.

Now, let an i_r-increasing and i_b-decreasing move be given. This is a good move, because it increases our imbalance with certainty. C_i denotes the change of the cost function after applying this move to the partition π. By an analogous argumentation we get a new random variable C_0' with the same distribution as C_0 and therefore with the same expectation. It can be easily seen that the following inequality is true

$$E(a(C_i)) - E(a(C_0)) \geq 0.$$

Now we look at the proposal and acceptance probabilities. u denotes the probability that the given partition π has got imbalance $i_\pi + 1$ after one transition. Thus u is a random variable on the set of all graphs. d denotes the corresponding probability that the given partition π has imbalance $i_\pi - 1$ after one step.

By combining the bounds of the proposal and the acceptance probabilities we get a short calculation

$$\begin{aligned} E(u) &\geq \frac{9b(n/3 - r)}{2n^2}E(a(C_0)) + \frac{9g(n/3 - r)}{2n^2}E(a(C_0)) \\ &= \left(\frac{2}{9} - \frac{2i_r}{n} + \frac{9i_r^2}{2n^2}\right)E(a(C_0)) \end{aligned}$$

and

$$\begin{aligned} E(d) &\leq \frac{9r(n/3-b)}{2n^2}\left(E(a(C_0)) - \frac{1}{20}\min\left\{\frac{i_r-i_b}{T_t},1\right\}\right) \\ &\quad + \frac{9r(n/3-g)}{2n^2}\left(E(a(C_0)) - \frac{1}{20}\min\left\{\frac{i_r-i_g}{T_t},1\right\}\right) \\ &= \left(\frac{2}{9}+\frac{5i_r}{2n}+\frac{9i_r^2}{2n^2}\right)E(a(C_0)) \\ &\quad - \frac{9r}{40n^2}\left((n/3-b)\min\left\{\frac{i_r-i_b}{T_t},1\right\} + (n/3-g)\min\left\{\frac{i_r-i_g}{T_t},1\right\}\right). \end{aligned}$$

Using $i_r = \max\{i_b, i_g, i_r\}$ and $i_b + i_g + i_r = 0$, we get $\exists c_1, c_2 \in \mathbb{R}_+$

$$E(u) \geq 2/9E(a(C_0)) - c_1\frac{i_r}{n}$$

and

$$E(d) \leq 2/9E(a(C_0)) + c_1\frac{i_r}{n} - c_2\min\left\{\frac{i_r}{T_t},1\right\}.$$

Because C_0 is symmetrically distributed, we obtain $E(a(C_0)) = 1/2$ and the proposition follows. □

Proof of Lemma 3:
We use the method of bounded differences (Theorem 15) applied to the set of indicator variables $e_{a,b} \in \{0,1\}$ of edges $(a,b) \in E$. Obviously, we can depict the random variables u and d as a function of the variables $e_{a,b}$. Varying one $e_{a,b}$ while keeping all others fixed, the proposal probability of a transition step (interchanging the positions of c and e) does not change. The acceptance function changes only for at most $2n$ possible values of (c,e) with $\{c,e\} \cap \{a,b\} \neq \emptyset$, because otherwise the edge (a,b) is not considered when deciding about the acceptance of one move. The cost change is at most one. A standard analysis argument yields that the change of the acceptance function is bounded by the derivative at an intermediate point times the change of the arguments. Thus, the change of the acceptance probability is bounded by $1/T_t$. By defining $c_{a,b} = 9/nT_t$ we get $\sum c_{a,b}^2 \leq 81/T_t^2$, and by applying the method of bounded differences (Theorem 15) the lemma follows. □

Proof of Proposition 4:
The proof idea is again to bound the sum of differences of

$$\sum_{\pi'} |P(\pi, \pi') - EP(\pi, \pi')|$$

and its expected value using the method of bounded differences, and to estimate the expected value.

Let C denote the cost change that we get by considering a possible transition from π to π' with $\pi \neq \pi'$. Jensen's inequality (see Proposition 16) and the concavity of the square root yields

$$\begin{aligned} E|P(\pi, \pi') - EP(\pi, \pi')| &= E\sqrt{(P(\pi, \pi') - EP(\pi, \pi'))^2} \\ &\leq \sqrt{E(P(\pi, \pi') - EP(\pi, \pi'))^2} \\ &= \frac{9}{2n^2}\sqrt{E(a(C) - E(a(C)))^2} \end{aligned}$$

Because $y = E(x)$ maximizes $E(x - y)^2$, we obtain

$$\begin{aligned} E|P(\pi, \pi') - EP(\pi, \pi')| &\leq \frac{9}{2n^2}\sqrt{E(a(C) - a(E(C)))^2} \\ &\leq \frac{9}{2n^2}\frac{1}{T_t}\sqrt{E(C - E(C))^2} \\ &= \frac{9}{2n^2 T_t}\sigma(C). \end{aligned}$$

In the last estimation we have used the fact that the differential quotient is bounded by the derivative. Since $\sigma(C) = O(\sqrt{n})$ (see the proof of Proposition 1), it follows that

$$\begin{aligned} E\sum_{\pi'} |P(\pi, \pi') - EP(\pi, \pi')| &= \sum_{\pi'} E|P(\pi, \pi') - EP(\pi, \pi')| \\ &\leq 2\sum_{\pi \neq \pi'} E|P(\pi, \pi') - EP(\pi, \pi')| \\ &= O\left(\frac{\sqrt{n}}{T_t}\right). \end{aligned} \tag{6}$$

Next we observe that we can depict $\sum_{\pi'} |P(\pi, \pi') - EP(\pi, \pi')|$ as a function of the indicator variables of the edges $e_{u,v}$. Varying one indicator variable the value of $EP(\pi, \pi')$ does not change, and $P(\pi, \pi')$ changes only for the $\frac{2n}{9}$ partitions π', whose corresponding transition is affected by the changed edge. The value of the change is bounded by at most $\frac{9}{2n^2 T_t}$ for $\pi \neq \pi'$. Therefore, $P(\pi, \pi')$ is changed by at most $\frac{1}{nT_t}$. It follows that

a change of $e_{u,v}$ causes a change of $\sum_{\pi'} |P(\pi,\pi') - EP(\pi,\pi')|$ of at most $\frac{1}{2nT_t} = c_{u,v}$. With $\sum c_{u,v}^2 \leq \frac{2}{T_t}$ and the method of bounded differences (Proposition 15) we get

$$P\left(|\sum_{\pi'} |P(\pi,\pi') - EP(\pi,\pi')| - E\sum_{\pi'} |P(\pi,\pi') - EP(\pi,\pi')|| > \frac{c\sqrt{n}}{T_t}\right)$$
$$\leq \quad 2e^{-c^2 n} \qquad (7$$

for arbitrary $c > 0$. Combining the Inequalities (6) and (7) we get

$$P\left(\sum_{\pi'} |P(\pi,\pi') - EP(\pi,\pi')| \geq O\left(\frac{\sqrt{n}}{T_t} + \frac{c\sqrt{n}}{T_t}\right)\right) \leq 2e^{-c^2 n}.$$

With Stirling's formula we obtain

$$\binom{n}{n/3} = e^{O(n)},$$

and thus, by choosing a sufficiently large c, the proposition follows. □

Proof of Theorem 2:
First of all we consider only the graphs $G \in \mathcal{G}$ with

$$\sum_{\pi'} |P(\pi,\pi') - EP(\pi,\pi')| \leq \frac{c\sqrt{n}}{T} \qquad (8)$$

for a suitable constant c. According to Proposition 4 every graph has this property with high probability for all partitions π.

The idea of this proof is to introduce a second process that should behave similarily as the current Simulated Annealing process, but without knowledge of the graph $G \in \mathcal{G}$.

Let a current partition π be given. We divide the interval $[0,1]$ into disjoint intervals of length $EP(\pi,\pi')$ (observe $\sum_{\pi'} EP(\pi,\pi') = 1$). The transitions of the second process are defined as follows: Generate a number uniformly at random in $[0,1]$. If it is in the interval $EP(\pi,\pi')$, our partition of the next step will be π'. In addition to that, a tape is given with points of time j and moves m_j. These moves are performed at time j instead of the transitions, which are defined by the random numbers.

Now we look at our basic process. Let a graph $G \in \mathcal{G}$ be given. We try to adapt the $P(\pi, \pi')$ intervals to the $EP(\pi, \pi')$ intervals beginning from left to right. We superimpose each $EP(\pi, \pi')$ interval with the corresponding $P(\pi, \pi')$ interval, matching up the left ends. If $EP(\pi, \pi') > P(\pi, \pi')$, we add the free space to a reserved space, which is used later. If $EP(\pi, \pi') < P(\pi, \pi')$, we put the missing part of the $P(\pi, \pi')$ interval on a stack. After matching all intervals we empty our stack by filling up the reserved space from left to right. Because both values sum up to one, this procedure fills the $[0, 1]$ interval completely. Obviously, we get a process that has the same distribution as our basic SA-process by generating a random number in $[0, 1]$ and choosing a partition π' in case of hitting parts of the interval $P(\pi, \pi')$.

The next step is to compare the transition probabilities of the basic and the second process. Beginning in the same initial state, the second process chooses a different partition in the next step, if and only if the chosen random number hits the reserved space, which is due to Lemma 4 of size $O(\sqrt{n}/T)$. Therefore we get

$$P(\text{basic and second process choose a different partition}) = O(\sqrt{n}/T).$$

Now we look at the probability of wrong moves (different from the basic process) of the second process at t subsequent steps. This number of wrong moves is stochastically dominated by $B(t, O(\sqrt{n}/T))$. Applying the inequality of Angluin and Valiant (Theorem 13) we can deduce:

The probability for the number of wrong moves to be $O(t\sqrt{n}/T)$ is $1 - \exp(-\Omega(t\sqrt{n}/T))$.

A short calculation (application of Stirling's formula) yields

$$\binom{t}{O(t\sqrt{n}/T)} = \exp(O(t\sqrt{n}\log t/T))$$

as a bound for the number of possibilities to distribute $O(t\sqrt{n}/T)$ wrong transitions in a set of t subsequent moves. Obviously, the basic process can perform only $O(n^2)$ different moves in each step. After fixing the $O(t\sqrt{n}/T)$ points of time of wrong moves, there are

$$(n^2)^{O(t\sqrt{n}/T)} = \exp(O(t\sqrt{n}\log n/T))$$

possibilities for the basic process to act at these points. Therefore, the number of different tapes necessary to ensure that there exists a tape, which can correct the second process (getting the same transitions of the basic and the second process), is bounded by

$$\exp(O(t\sqrt{n}\log n/T)$$

(we assume $t = O(poly(n))$.

Now, let a sequence of t uniformly at random generated numbers in $[0,1]$ be given. We can be sure, that the second process, started with all possible tapes described above, yields a set S of sequences of partitions, which contains the sequence of partitions of the basic process. This set of sequences contains $\exp(O(t\sqrt{n}\log n/T)$ states.

Applying Lemma 3 concerning the deviation we obtain that the probability over all graphs G for some partition in the above described set to be α-deviant, is bounded by

$$\exp(t\sqrt{n}\log n/T - O(\alpha^2 T^2)),$$

and the theorem follows by the choice of t. □

Proof of Proposition 5:
Proof: First, we define a new Markov chain $(X'_t)_{t\in\mathbb{N}}$ on the natural numbers. Let $\Omega_k = \{\omega \in \Omega | \Xi(\omega) = k\}$ be a partition of the state space. We define the time inhomogeneous transition probabilities $P_t^{X'}$ of $(X'_t)_{t\in\mathbb{N}}$ as follows

$$\Xi(X_0) = X'_0 \quad \text{and} \quad P_t^{X'}(i,j) = P(X_{t+1} \in \Omega_j | X_t \in \Omega_i).$$

One gets easily by induction

$$P(X_t \in \Omega_i) = P(X'_t = i).$$

Thus, $\Xi(X_t)$ and X'_t have the same distribution. Now we compare X'_t with ξ_t. Applying Theorem 5.8 of LINDVALL [14] we can deduce

If two Markov kernels K^Y and K^Z exist on $\mathbb{R}$ with

$$K^Y(x,[y,\infty)) \leq K^Z(x',[y,\infty))$$

for all $x, x', y \in \mathbb{R}$ with $x \leq x'$, then there exist two Markov chains Y and Z governed by K^Y and K^Z with $Y_t \leq Z_t$ for all t, if the starting distribution of Y stochastically dominates the starting distribution of Z.

Identifying ξ_t with the transition kernel

$$\begin{aligned} K_t^Y(.,.) : \mathbb{N} \times 2^{\mathbb{N}} &\to [0,1] \\ (x,A) &\mapsto P(\xi_t \in A | \xi_{t-1} = x) \end{aligned}$$

and X'_t with the transition kernel

$$\begin{aligned} K_t^Z(.,.) : \mathbb{N} \times 2^{\mathbb{N}} &\rightarrow [0,1] \\ (x, A) &\mapsto P(X'_t \in A | X'_{t-1} = x), \end{aligned}$$

by the dominance properties in the proposition it is easily seen that

$$K^Y(x, [y, \infty)) \leq K^Z(x', [y, \infty))$$

is satisfied for all $x, x', y \in \mathbb{R}$ with $x \leq x'$. Due to $X'_0 = \Xi(X_0) \geq \xi_0$ the stochastical dominance of the starting distribution is also fulfilled, and the proposition follows. □

Proof of Lemma 6:
According to FELLER [6] we get for the expected first hitting time D_i of $n^{1/3}$, starting lb_t^b at an arbitrary $i \in \{0, \ldots, n^{1/3}\}$

$$D_i = \frac{i}{2\alpha} + A + B \left(\frac{k + 2\alpha}{k - 2\alpha} \right)^i$$

with $A, B \in \mathbb{R}$ that must fit the boundary conditions

$$D_0 = (k + 2\alpha) D_1 + (1 - k + 2\alpha) D_0 \quad \text{and} \quad D_{n^{1/3}} = 0.$$

Solving for A and B yields

$$\begin{aligned} D_0 &= -\frac{n^{1/3}}{4\alpha} + \frac{k + 2\alpha}{16\alpha^2} \left[\left(\frac{k + 2\alpha}{k - 2\alpha} \right)^{n^{1/3}} - 1 \right] \\ &= O(n^{2/3}) \end{aligned}$$

for the chosen parameter values α and r. Let t_f be the random variable denoting the first hitting time of $n^{1/3}$ of the process lb_t^b.

With the Markov inequality we get

$$P(t_f > n^{2/3+\epsilon}) \leq \frac{E(t_f)}{n^{2/3+\epsilon}} = \frac{D_0}{n^{2/3+\epsilon}} = O(n^{-\epsilon})$$

and the lemma follows. □

Proof of Lemma 8:
According to FELLER [6] we get the following upper bound for the probability that lb_t^j reaches $\frac{n^{1/3+j\gamma}}{2}$ earlier than $n^{1/3+(j+1)\gamma}$, starting at $n^{1/3+j\gamma}$ (observe that lb_t^j is a random walk without reflecting barriers)

$$\frac{q^{\frac{1}{2}n^{1/3+j\gamma}} - q^{n^{1/3+(j+1)\gamma}-n^{1/3+j\gamma}}}{1 - q^{n^{1/3+(j+1)\gamma}-n^{1/3+j\gamma}}} = e^{-n^{\Omega(1)}}$$

for

$$q = \frac{k - \Theta(n^{-1/3})}{k - \Theta(n^{-1/3+(j+1)\gamma})}$$

being the ratio of up- and downwards probabilities of lb_t^j.

The expected time to hit one of the barriers $\frac{n^{1/3+j\gamma}}{2}$ or $n^{1/3+(j+1)\gamma}$ can be bounded with the help of an analogous argument as in Lemma 6 by $O(n^{2/3})$. Applying the Markov inequality yields the desired result. □

Proof of Lemma 11:
The proof idea is to estimate the probability to choose a specific vertex in the sampling process and to use the method of bounded differences to bound the deviation from the expected values.

We note first that the size of the sample is less than the guaranteed imbalance. This is important, because due to this fact we can compare our actual sampling process with a sequence of independent Bernoulli trials. Let again red be the most frequent color in the smaller set of the partition π, and S_r be the number of red vertices in the sample. Obviously,

$$p_r^u = \frac{n/9 + i_r}{n/3 - n^{3/4-\epsilon}}$$

is an upper bound and

$$p_r^l = \frac{n/9 + i_r - n^{3/4-\epsilon}}{n/3}$$

a lower bound for the probability to choose a red vertex during the sampling process. Thus, $p_r^u n^{3/4-\epsilon}$ is an upper bound and $p_r^l n^{3/4-\epsilon}$ is a lower bound for the expected value $E(S_r)$. Following the same argument we get

$$\forall k \in \mathbb{N} \quad P(S_r \geq k) \leq P(B(n^{3/4-\epsilon}, p_r^u) \geq k)$$

and

$$\forall k \in \mathbb{N} \quad P(S_r \leq k) \leq P(B(n^{3/4-\epsilon}, p_r^l) \leq k).$$

Applying the method of bounded differences with $\delta = n^{5/12}$ and the sample space of $n^{3/4-\epsilon}$ events, we obtain:

$$P(|B(n^{3/4-\epsilon}, p_r^u) - n^{3/4-\epsilon} p_r^u| > \delta) \leq 2e^{\frac{-2\delta^2}{n^{3/4-\epsilon}}} \leq 2e^{n^{-1/6+\epsilon}}.$$

We get for the difference of the expectations

$$n^{3/4-\epsilon}(p_r^u - p_r^l) \leq O(n^{1/2-2\epsilon}),$$

and, by considering the bound $\delta = n^{5/12} \leq n^{1/2-\epsilon}$, the result follows.

By an analogous argument we get the same result for the other colors. □

References

[1] Alon, N.; Kahale, N.: *A spectral technique for coloring random 3-colorable graphs*, Proceedings of the 26th Symposium on Theory of Computing, 1994

[2] Angluin, D.; Valiant, L.G.: *Fast probabilistic algorithms for Hamiltonian circuits and matchings*, Journal of Computer System Science 18, 1979

[3] Blum, A.; Spencer, J.: *Coloring Random and Semi-Random k-Colorable graphs*, Journal of Algorithms 19, 1995

[4] Chung, K.L.: *Markov chains with stationary transition probabilities*, Springer Verlag, Heidelberg, 1960.

[5] Dyer, M.; Frieze, A.: *The Solution of Some Random NP-Hard Problems in Polynomial Expected Time*, Journal of Algorithms 10, 1989

[6] Feller, W.: *An introduction to probability theory and its applications, Volume 1*, John Wiley & Sons, New York, 1950

[7] Garey, M.R.; Johnson, S.J.: *Computers and Intractability*, W.H. Freeman and Company, 1979

[8] Jensen, T.; Toft, B.: *Graph Coloring Problems*, John Wiley & Sons, 1995

[9] Jerrum, J; Sorkin, G.: *Simulated Annealing for Graph Bisection*,u Technical Report 1993, LFCS, University of Edinburgh

[10] Johnson, D.S.; Aragon, C.R.; McGeoch, L.A.; Schevon, C.: *Optimization by Simulated Annealing: An Experimental Evaluation*, Operations Research 39, 1991

[11] Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P.: *Optimization by Simulated Annealing*, Science 220, 1983.

[12] Kucera, L.: *Expected behaviour of graph coloring algorithms*, Lecture Notes in Computer Science 56, 1977

[13] Laarhoven, P.J.M.; Aarts, E.H.L.: *Simulated Annealing: Theory and Applications*, Kluwer Academic Publishers, 1989.

[14] Lindvall, T.: *Lectures on the Coupling method*, John Wiley & Sons, 1992

[15] Metropolis, N.; Rosenbluth, M.; Rosenbluth, M.; Teller, A.; Teller, E.: *Equation of state calculations by fast computer machines*, Journal of Chemical Physics 21, 1953

[16] Nolte, A.; Schrader, R.: *Coloring in Sublinear Time*, Proceedings ESA '97, Lecture Notes in Computer Science 1284, Springer

[17] Petford, A.D.; Welsh, D.J.A.: *A Randomized 3-Coloring Algorithm*, Discrete Mathematics 74 (1989), North Holland

[18] Prömel, H.J.; Steger, A.: *Random l-colorable graphs* Random Structures and Algorithms 6, 1995

[19] Sasaki, G.H.; Hajek, B.: *The Time Complexity of Maximum Matching by Simulated Annealing*, Journal of the Association for Computing Machinery 35, 1988

[20] Turner, J.: *Almost All k-Colorable Graphs Are Easy to Color*, Journal of Algorithms 9, 1988