Coloring Random $K-$ Colorable Graphs

by

Peter Pengxiang Yan

B Sc , Fuzhou University, 1982

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Computer Science

We accept this thesis as conforming
to the required standard

[redacted]

Dr V King, Supervisor (Department of Computer Science)

[redacted]

Dr M Fellows, Departmental Member (Department of Computer Science)

[redacted]

Dr G MacGillivray, Outside Member (Department of Mathematics and Statistics)

[redacted]

Dr A Gupta, External Member (Simon Fraser University)

Examiners:

[redacted]

Dr. V. King, Supervisor (Department of Computer Science)

[redacted]

Dr. M. Fellows, Departmental Member (Department of Computer Science)

[redacted]

Dr. G. MacGillivray, Outside Member (Department of Mathematics and Statistics)

[redacted]

Dr. A. Gupta, External Member (Simon Fraser University)

Acknowledgements

I would like to take this opportunity to express my profound gratitude to my supervisor Dr. Valerie King for her excellent academic guidance and incessant encouragement throughout the preparation of this thesis. I also express my deep compliments to Dr. J. Ellis for his wonderful lecture and helpful discussion on this subject. Dr. Nabil Kahale also receives my compliments for his kindly helps.

I also wish to acknowledge my gratitude to the University of Victoria and the Department of Computer Science for the financial assistance and the excellent research environment.

CONTENTS

Chapter 1

Introduction

Since Cook introduced the foundations for the theory of NP-complete in 1971, it has become one of the most active research areas in the theory of computation. Many computing problems have been proved to be NP-complete by various researchers [15]. Because it is impossible to solve an $NP-$complete problem in polynomial time unless $P = NP$, finding an approximate solution for $NP-$ complete problem then becomes important. In this thesis, we will consider the graph coloring problem which is one of the well-known NP-complete problems.

§1.1 Background

Given a simple undirected graph G we denote V as its vertex set and E as the edge set. A vertex coloring of a graph G is proper if no adjacent vertices receive the same color. The chromatic number $\chi(G)$ of G is the minimum number of colors in a proper vertex coloring of G. In this thesis, the graph G is assumed to be k-colorable with $k \geq 3$ a fixed positive integer. The graph coloring problem is to determine the minimum number of colors needed to color G. Graph coloring has many applications in scheduling and timetabling.

Graph coloring is one of the most notorious of the NP-complete problems. In 1972, Karp [22] first showed that the problem is NP-complete. Stockmeyer [29] strengthened this by showing that it remains NP-complete for fixed $k \geq 3$. Those results have led many researchers to seek approximation algorithms capable of producing colorings that do not use too many extra colors. Garey and Johnson [15] proved that unless $P = NP$, no polynomial time approximation algorithm can guarantee the use of fewer than $2\chi(G)$ colors. Furthermore, Johnson [20] showed that for many popular heuristics, there are 3-colorable graphs on n vertices for which the heuristics require $O(n)$ colors. It is therefore unlikely that there are efficient algorithms for coloring optimally an arbitrary k-chromatic input graph with $k \geq 3$ and researchers should address the approximate solution issue.

Numerous algorithms for finding approximate solutions have been proposed by various researchers The analysis of approximation algorithms for graph coloring started with the work of Johnson [20] who shows that a version of the greedy algorithm gives an $O(\frac{n}{\log n})$-approximation algorithm for k-coloring Wigderson [31] improved this bound by giving an elegant algorithm which uses $O(n^{\frac{1}{k-1}})$ colors to properly color a k colorable graph Subsequently other polynomial time algorithms were provided by Blum [5] which use $O(n^{\frac{3}{8}} \log^{\frac{8}{5}} n)$ colors to properly color an n-vertex 3-colorable graph This result generalizes to coloring a k-colorable graph with $O(n^{1-\frac{1}{k-\frac{4}{3}}})$ colors His idea is based on examing second-order neighborhoods, i e neighborhoods of the neighborhoods of vertices, rather than just immediate neighbors of vertices as in previous approaches The best known performance guarantee for general graphs is due to Halldorsson [18] who provided a polynomial time algorithm using a number of colors which is within a factor of $O(n(\log\log n)^2/\log^3 n)$ of the optimum

The disappointing results of deterministic algorithms and analysis for graph coloring suggests that probabilistic methods may provide a more effective way of proposing new algorithms and evaluating candidate algorithms Randomized algorithms hence are introduced to obtain better solutions We define a randomized algorithm as an algorithm which is allowed access to a source of independent, unbiased random bits it is then permitted to use these random bits to influence its computation.

Blum's result was improved by Karger, Motwani and Sudan [21] recently. They presented a randomized polynomial time algorithm which colors a 3-colorable graph on n vertices with $min\{O(\check{d}^{\frac{1}{3}} \log^{\frac{4}{3}} \check{d}),\ O(n^{\frac{1}{4}} \log n)\}$ colors where $\check{d}$ is the maximum degree of any vertex. It is the first time that the number of colors used in the algorithm is associated with the maximum degree. They used algorithms for semidefinite programming to obtain an optimum solution to a relaxed version of the problem, and a randomized strategy for "rounding" this solution to a feasible but approximate solution to the original problem. Mahajan and Ramesh [32] derandomized the algorithm of Karger, Motwani and Sudan and obtained a polynomial time deterministic algorithm with the same approximation ratio.

Rather than evaluating an algorithm based on its performance in the worst case (for any graph), one may ask how well the algorithm performs on a random graph. Here we say that a graph is a random graph if we specify certain probabilistic properties on its vertex number and edge distribution. Some random graph classes used frequently by many researchers are: (1) $G(n,p)$: $n-$vertex graph which each pair of vertices is joined by an edge with a prescribed probability p. (2) $G(n, p, 3)$: A class of $3-$colorable random graphs, constructed in this way: given n vertices, each vertex first picks a color randomly, independently and uniformly, among the three possibilities, and next every pair of vertices of distinct colors becomes an edge with probability p Many deterministic algorithms have been shown to work very well on random graph models

One of the well-known algorithms designed for random graphs is a greedy algorithm proposed by Grimmett and McDiarmid[15] which we will discuss later Their algorithm produces very good results for random graphs generated as in (1) above even though it is extremely simple This result suggests that the usual model is unrealistic, since it makes even the most simple-minded algorithm look good In order to obtain meaningful comparative information we should try to select a more difficult probability distribution, one that poses some challenges for candidate algorithms to overcome

Turner [30] shows that many random graphs can be colored easily Most randomized algorithms are applicable on dense graphs, which are relatively easy to color To color "sparse" graphs, Alon and Kahale [2] recently proposed a new algorithm which gives an optimal proper coloring for a 3−colorable graph (random graph model) almost surely (Here and in what follows, *almost surely* always means with probability that approaches 1 as n tends to infinity and *a constant* means a number independent of n where n is the vertex number of the graph G) They employed spectral techniques

Since there is an one to one relationship between a graph and its adjacency matrix and many properties of a matrix are associated with its eigenvalues and eigenvectors, it is natural to approach the coloring problem by investigating the relationship between the graph coloring and matrix spectral properties We will show this relationship can be approximated by a critical function which we define in the next chapter

This thesis which studies the random graph class $G(n, p, k)$ that we define later is organized as follows In Chapter 2 we will introduce a critical function, which has a clear geometric explanation of the graph coloring problem and produces a good approximate solution which will be used in Chapter 3 It is hoped that this result will indicate a new approach for coloring more general random graphs

In Chapter 3 of this thesis we will propose a simpler way to generate the initial solution in Alon and Kahale's algorithm Chapter 4 gives the correctness proof of the algorithm

The main idea of the algorithm can be described as follows We first map all n vertices to the real $k-1$ dimensional space R^{k-1} such that points in the same color class are put together "closely" with small exceptions We will show that this map can be obtained by "rotating eigenvectors", as is motivated by Alon and Kahale's paper [2] Then we get a good approximation of the coloring such that all vertices are colored properly except a relatively small number of vertices By the fact that almost all neighbors of a vertex are properly colored, we can reduce the number of wrongly colored vertices by half in each iteration Finally we can show that only a

$O(\log_k n)$ number of vertices remain uncolored, therefore, we can use an exhaustive technique to color all remaining vertices. This algorithm takes polynomial time and produces a proper coloring almost surely.

§1.2 Preliminaries

In this section, we introduce some definitions and known results that will be used in the following chapters.

Definition 1 *Given a simple undirected graph $G = (V, E)$, where V is the set of vertices, and E is the set of edges, a k-coloring of G is an assignment $C : V \mapsto \{1, \ldots, k\}$ such that if $C(v) = C(w)$ then edge (v, w) is not in E. The chromatic number $\chi(G)$ is the minimum k for which there exists a k-coloring of G.*

In this thesis, we will consider a random k-colorable graph on a set of n vertices, which is generated as follows. First, split the vertices arbitrarily into k equal color classes and then choose every pair of vertices of distinct color classes, randomly and independently to be an edge with probability p. $p > \frac{c}{n}$, $c >> 1$. Here we assume that n is a multiple of k. We denote this class of random graphs by $G(n, p, k)$. Let G' be the graph obtained from G by deleting all edges incident to vertices of degree greater than $(2k - 1)d$, $d = \frac{np}{k}$.

Definition 2 *An independent set is $V' \subseteq V$ such that for $v,\ w \in V'$, $(v,\ w) \notin E$. A clique is a set $V' \subseteq V$ such that for $v,\ u \in V'$, $(v, u) \in E$. The complement $\overline{G}$ of G is the graph in which each pair $v,\ w$ is an edge in $\overline{G}$ exactly when it is not an edge in G.*

Therefore, a $k-$coloring partitions the vertex set into independent sets (or color classes), with each color class being the set of vertices of a given color. An independent set in G is a clique in $\overline{G}$.

Definition 3 *For any $x = (x_i)_1^n \in R^n$ and $y = (y_i)_1^n \in R^n$, let $< x, y > = \sum_{i=1}^n x_i y_i$, $\|x\|^2 = < x, x >$. x^t is the transposition of x.*

Denote the Rayleigh quotient as

$$R(x) = \frac{< x, Ax >}{< x, x >},$$

and denote the probability of a random variable X by $P(X)$, the expectation by $E(X)$, and the variance by $var(X)$.

Definition 4 *Let $f(n)$, $g(n)$: $R \rightarrow R$ be two non-negative real valued functions.*

1. *We say that $f(n) = O(g(n))$ if there exist positive number c and N such that, for all $n \geq N$, $f(n) \leq cg(n)$.*

2. *We say that* $f(n) = \Omega(g(n))$ *if there exist positive number* c *and* N *such that, for all* $n \geq N$, $f(n) \leq cg(n)$

3. *We say that* $f(n) = \Theta(g(n))$ *if* $f(n) = O(g(n))$ *and* $f(n) = \Omega(g(n))$

4. *We say that* $f(n) = o(g(n))$ *if* $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$

The following well known theorems will be used extensively in the analysis of our algorithm.

Theorem 1 *(Courant-Fischer Theorem [26], pp. 116) Given any linearly independent vectors* $a_1, a_2, \ldots, a_r$ *in* R^n, *let* R^{n-r} *be the subspace consisting of all vectors in* R^n *which are orthogonal to* $a_1, a_2, \ldots, a_r$. *Let* A *be a real symmetric matrix and denote eigenvalues of* A *by* $\lambda_n(A) \geq \lambda_{n-1}(A) \geq \ldots \geq \lambda_1(A)$. *Then*

$$\lambda_{n-r}(A) = \min_{R^{n-r}} \max_{x \in R^{n-r}} R(x), \quad r = 1, 2, \ldots, n$$

Theorem 2 *(Chebyshev's Inequality [28], pp. 61) Let* X *be a random variable with expectation* μ_X *and standard deviation* σ_X. *Then for any* $t \in R^+$

$$P[|X - \mu_X| > t\sigma_X] \leq \frac{1}{t^2}$$

Theorem 3 *(Chernoff Bound Theorem [28], pp. 84) Let* $X_1, X_2, \ldots, X_n$ *be independent Bernoulli trails with* $P[X_i = 1] = p_i$, $p_i \in (0, 1)$. *Let* $X = \sum_{i=1}^n X_i$ *and* $\mu = \sum_{i=1}^n p_i > 0$. *Then for any* $\delta > 0$

$$P[X > (1+\delta)\mu] < [\frac{e^{\delta}}{(1+\delta)^{1+\delta}}]^{\mu}$$

Next we give some probabilistic results about the graph G.

Lemma 1 *The expectation of each vertex degree of G is $(k-1)d$. The variance of each vertex degree of G is $(k-1)(1-p)d$.*

Proof Let x be the random variable which is equal to 1 if there is an edge from a vertex to any vertex in a different color class and 0 otherwise. Then x satisfies the Bernoulli distribution with expectation equal to p. Let X be the random variable defined to be the degree of a vertex. Then X is the sum of $(k-1)\frac{n}{k}$ random variables x, which satisfy the binomial distribution. So the expectation of X is $(k-1)\frac{n}{k}p = (k-1)d$ ([28], pp. 400). The variance can be proven similarly, completing the proof. □

Lemma 2 *For any constant $\beta > 0$ such that $2^{-\beta d}n \geq 1$, almost surely for any subset X of V of $2^{-\beta d}n$ vertices, $e(X, V) \leq (2k-1)d|X|$, where $e(X, V)$ is the number of edges (u, v), with $u \in X$.*

Proof Let x be the random variable defined to be the edge number from X to V. Then the expectation of x, $\mu = (k-1)\frac{np}{k}|X| = (k-1)d|X| = (k-1)d2^{-\beta d}n$

Since $d = \frac{np}{k}$, if $|X| = 2^{-\beta d} n$ is bounded by a constant, then $\lim_{n \to \infty} d = \infty$, if $|X|$ is unbounded, i.e., $\lim_{n \to \infty} |X| = \infty$. Therefore $\mu \to \infty$ as $n \to \infty$. Furthermore

$$\begin{aligned} & P(x > (2k-1)d|X|) \\ & = P(x > (1 + \tfrac{k}{k-1})(k-1)d|X|) \\ & = P(x > (1 + \tfrac{k}{k-1})\mu) \end{aligned} \tag{1.1}$$

Set $\delta = \frac{k}{k-1}$. By using Chernoff's Bound Theorem, we have

$$\begin{aligned} & P(x > (2k-1)d|X|) \\ & < [\tfrac{e^{\delta}}{(1+\delta)^{1+\delta}}]^{\mu} \\ & = [e^{\delta-(1+\delta)\log(1+\delta)}]^{\mu} \\ & = e^{(\delta-(1+\delta)\log(1+\delta))\mu} \end{aligned} \tag{1.2}$$

Set $f(\delta) = \delta - (1+\delta)\log(1+\delta)$. Since $f(0) = 0$ and $f' = -\log(1+\delta) < 0$, we have $f(\delta) < 0$ for any $\delta > 0$. Therefore, $\lim_{n \to \infty} P(x > (2k-1)d|X|) = 0$, i.e., $x \leq (2k-1)d|X|$ almost surely. We have proven the lemma. $\square$

By applying the above lemma, we have

Lemma 3 *Let E' be the edge set of G', then $|E - E'| = 2^{-\Omega(d)} n$ almost surely.*

§1.3 Algorithm Classes

With the help of last section, we can classify graph coloring algorithms into the following categories

Greedy [12], [17] and [?] Find large independent sets and color each of them with one color. One of the simplest implementations is the greedy algorithm introduced by Grimmett and McDiarmid [17], which can be described as follows: let G be any graph with vertex set $\{v_1, v_2, \ldots, v_n\}$. Color v_1 with color 1, and then proceed to color the remaining vertices in increasing order, using color j to color v_i $(2 \leq i \leq n)$ where j is the least positive integer such that no vertex already colored by color j is joined (in G) to v_i. For the random graph $G(n, p)$ introduced in last section, Grimmett and McDiarmid showed that $\chi(G)$ is at least $\frac{1}{2} \log \frac{1}{1-p} \frac{n}{\log n} + o(\frac{n}{\log n})$, and the number of colors required by the algorithm is $\frac{n}{\log n} \log \frac{1}{1-p}$ almost surely.

Partition [5] Partition the vertices by some means, and then attempt to remove conflicts by moving vertices from one partition to another. These methods first produce an approximately correct coloring in which some conflicts may remain, i.e., the same colors are assigned to both endpoints of an edge. The conflicts are then resolved by assigning new colors to the vertices.

Clique [30] After choosing the first vertex, choose vertices with a maximum number of constraints on the colors available to them. This is almost the opposite of the first method as now we choose vertices that form large cliques with respect to the vertices already chosen. Turner's algorithm attempts to find a $k-$ coloring of a graph $G = (V, E)$, where k is assumed to be an input parameter. To illustrate the algorithm, define a partial coloring of G to be a mapping $c : V \rightarrow [0, n]$. The algorithm starts by constructing the partial coloring defined by $c(x) = 0$ for all $x \in V$ and then attempt to convert this to a complete proper coloring. For each vertex x, set $avail(x) = \{i | 1 \leq i \leq n$ and if $(x, y) \in E$, $c(y) \neq i\}$. This algorithm has two phases. In the first phase it attempts to find a $k-$clique by repeating the following step k times.

Clique Finding Step: Select a vertex x adjacent to all previously selected vertices.

If the algorithm finds a $k-$clique, it colors each of the vertices in the clique with a distinct color in $[1, k]$ and starts the second phase which consists of repeated applications of the following rule.

Coloring Rule: Select an uncolored vertex x for which $|avail(x) \cap [1, k]| = 1$ and $c(x) = \min avail(x)$.

If neither rule can be applied before completion, the algorithm fails.

Zykov [13] Ellis and Lepolesa's algorithm consists of the following steps. First construct a Zykov's tree as follows: suppose there exist two non adjacent vertices u and v in G. Let G_1 denote the graph obtained from G by adding an edge $\{u, v\}$ and let G_2 denote the graph obtained by identifying u and v, i.e., replacing u and v by a single vertex that is adjacent to each vertex which was adjacent to either u or v. Repeatedly apply these constructions until both G_1 and G_2 are cliques. Then do partial traversals of the Zykov's tree as follows: (1) The procedure can, if necessary, search the entire tree. If a clique is reached which is too large, the procedure does not terminate, it backtracks and continues. (2) At each branch of the tree, formed either by adding an edge or by identifying vertices, a test is made for the presence of a $(k+1)-$clique. The test procedure is fast because it is restricted to the one portion of the graph at which the edge additions and vertex identifications are taking place. Consequently a "No" answer may be incorrect, i.e., there may exist a $(k+1)-$clique, somewhere in the graph, but the test missed it. Hence it may allow unnecessary exploration to continue. The purpose of the test is twofold. Firstly it has a tendency to prune the search tree. Secondly, it ensures that the procedure terminates correctly on graphs that are not $k-$ colorable.

Spectral Technique [2] and [21] Derive approximation algorithms based on the spectral properties of the adjacency matrix of the given graph. This thesis generalizes Alon and Kahale's result

Chapter 2

Critical function

In this chapter, we will find an approximate solution for coloring a k-colorable graph. We first show the following existence lemma.

Lemma 4 *For $k \geq 3$, there exists a series $\{u_i^k\}_{i=1}^k$, $u_i^k = (x_1^{i,k}, x_2^{i,k}, \ldots, x_{k-1}^{i,k})$, such that $\|u_i^k\|^2 = \frac{k-1}{2k}$, $\|u_i^k - u_j^k\| = 1$, $i \neq j$, $i,j = 1, \ldots, k$. If we set $U_j^k = (x_j^{1,k}, x_j^{2,k}, \ldots, x_j^{k,k})$, then we have $(U_i^k, U_j^k) = 0$, $i \neq j$, $\sum_{l=1}^k x_i^{l,k} = 0$, $\|U_j^k\|^2 = \frac{1}{2}$, $i,j = 1, 2, \ldots, k-1$.*

Proof We prove the lemma by constructing the series as follows. For $k = 3$, $u_1^3 = (-\frac{1}{2}, -\frac{1}{\sqrt{12}})$, $u_2^3 = (\frac{1}{2}, -\frac{1}{\sqrt{12}})$, $u_3^3 = (0, \frac{2}{\sqrt{12}})$ is the required solution.

Suppose that for $k = m - 1$, the lemma is true, then let

$u_i^m = (x_1^{i,m-1}, \ldots, x_{m-2}^{i,m-1}, -\frac{1}{\sqrt{2m(m-1)}})$ for $i = 1, 2, \ldots, m-1$, and $u_m^m = (0, \ldots, 0, \frac{m-1}{\sqrt{2m(m-1)}})$

We claim $\{u_i^m\}_{i=1}^m$ satisfies the lemma. We show it in five parts.

1. $\|u_j^m\|^2 = \frac{m-1}{2m}$, $j = 1, 2, \ldots, m$. Obviously, $\|u_m^m\|^2 = \frac{m-1}{2m}$. For $j < m$,

$$\begin{aligned}\|u_j^m\|^2 &= \|u_j^{m-1}\|^2 + \left(\frac{1}{\sqrt{2m(m-1)}}\right)^2 \\ &= \frac{m-2}{2(m-1)} + \frac{1}{2m(m-1)} = \frac{m-1}{2m}\end{aligned}$$

2. $\|u_i^m - u_j^m\| = 1$ for $1 \le i < j \le m$. For $1 \le i < j < m$, $\|u_i^m - u_j^m\| = \|u_i^{m-1} - u_j^{m-1}\| = 1$. For $i < m$,

$$\begin{aligned}\|u_i^m - u_m^m\|^2 &= \|u_i^{m-1}\|^2 + \left(\frac{m}{\sqrt{2m(m-1)}}\right)^2 \\ &= \frac{m-2}{2(m-1)} + \frac{m}{2(m-1)} = 1\end{aligned}$$

3. $\sum_{l=1}^m x_i^{l,m} = 0$, $i = 1, \ldots, m-1$. For $i < m-1$, $\sum_{l=1}^m x_i^{l,m} = \sum_{l=1}^{m-1} x_i^{l,m-1} = 0$

and

$$\sum_{l=1}^m x_{m-1}^{l,m} = -\frac{m-1}{\sqrt{2m(m-1)}} + \frac{m-1}{\sqrt{2m(m-1)}} = 0$$

4. $(U_i^m, U_j^m) = 0$, $1 \le i < j \le m-1$.

For $1 \le i < j < m-1$, $(U_i^m, U_j^m) = (U_i^{m-1}, U_j^{m-1}) = 0$.

For $j < m-1$, $(U_j^m, U_{m-1}^m) = -\frac{1}{\sqrt{2m(m-1)}} \sum_{l=1}^{m-1} x_i^{l,m-1} = 0$.

5. $\|U_j^m\|^2 = \frac{1}{2}$, $1 \le j \le m-1$. For $1 \le j < m-1$, obviously $\|U_j^m\|^2 = \frac{1}{2}$. For $j = m-1$, we have $\|U_{m-1}^m\|^2 = (m-1)\frac{1}{2m(m-1)} + \frac{(m-1)^2}{2m(m-1)} = \frac{1}{2}$.

By combining 1–5, we prove the lemma. □

We will use these k points $\{u_1^k, \ldots, u_k^k\}$ in R^{k-1} to represent k colors. We call these k points color points. Then we can restate our coloring problem as follows. Given a $k-$ colorable graph $Q(V, E, k)$, where $V = \{v_1, \ldots, v_n\}$ is the vertex set and E is the edge set, find a vertex map $T : \{v_1, \ldots, v_n\} \mapsto \{u_1^k, \ldots, u_k^k\}$, such that $T(v_i)$, $i = 1, \ldots, n$ is a proper coloring. In this case, we say T is an optimal map. Denote $A' = (a'_{ij})$ the adjacency matrix of $Q(V, E, k)$. We introduce a discrete critical problem.

Critical Problem 1 *Let*

$$f(\tilde{\eta}_1, \ldots, \tilde{\eta}_n) = \sum_{1 \le i, j \le n} a'_{ij} \|\tilde{\eta}_i - \tilde{\eta}_j\|^2.$$

Find n points $\tilde{\eta}_j \in \{u_1^k, \ldots, u_k^k\}$, $j = 1, \ldots, n$ such that f reaches the maximum.

The relationship between coloring graph Q and Critical Problem 1 can be described as the following lemma.

Lemma 5 $f(\tilde{\eta}_1, \ldots, \tilde{\eta}_n) = \sum_{1 \le i, j \le n} a'_{ij}$ *if and only if, any two adjacent vertices are mapped to different color points, i.e., the map is an optimal map.*

Proof The lemma can be seen from the fact that $f(\tilde{\eta}_1, \ldots, \tilde{\eta}_n) \le \sum_{1 \le i, j \le n} a_{ij}$ since $\|\tilde{\eta}_i - \eta_j\| \le 1$. □

In Critical Problem 1, if we consider those n points in R^{k-1} as $k-1$ points in R^n, the critical function f can be rewritten as follows. Let $\eta_i = (\xi_1^i, \ldots, \xi_n^i)$, $i = 1, \ldots, k-1$, $\tilde{\eta}_j = (\xi_j^1, \ldots, \xi_j^{k-1})$, $j = 1, \ldots, n$ and

$$f(\eta_1, \ldots, \eta_{k-1}) = \sum_{1 \le i,j \le n} a'_{ij} \sum_{l=1}^{k-1} (\xi_i^l - \xi_j^l)^2$$

Therefore, we can give a simpler expression of f

Lemma 6

$$f(\eta_1, \ldots, \eta_{k-1}) = 2 \sum_{i=1}^{k-1} \eta_i B \eta_i^t,$$

where $B = D - A'$ and D is a diagonal matrix with entries $\sum_{j=1}^n a'_{ij}$ (degree of vertex i), $i = 1, \ldots, n$. Moreover, $B \ge 0$(semidefinite).

Proof Since A' is symmetric, we have

$$\begin{aligned}
f(\eta_1, \ldots, \eta_{k-1}) &= \textstyle\sum_{1 \le i,j \le n} a'_{ij} \sum_{l=1}^{k-1} (\xi_i^l - \xi_j^l)^2 \\
&= \textstyle\sum_{1 \le i,j \le n} a'_{ij} \sum_{l=1}^{k-1} ((\xi_i^l)^2 - 2\xi_i^l \xi_j^l + (\xi_j^l)^2) \\
&= \textstyle\sum_{l=1}^{k-1} \sum_{1 \le i,j \le n} a'_{ij} ((\xi_i^l)^2 - 2\xi_i^l \xi_j^l + (\xi_j^l)^2) \\
&= \textstyle\sum_{l=1}^{k-1} \sum_{1 \le i,j \le n} a'_{ij} (\xi_i^l)^2 + \sum_{l=1}^{k-1} \sum_{1 \le i,j \le n} a'_{ij} (\xi_j^l)^2 - 2 \sum_{l=1}^{k-1} \sum_{1 \le i,j \le n} a'_{ij} \xi_i^l \xi_j^l \\
&= \textstyle\sum_{i=1}^{k-1} \eta_i D \eta_i^t + \sum_{i=1}^{k-1} \eta_i D \eta_i^t - 2 \sum_{i=1}^{k-1} \eta_i A' \eta_i^t \\
&= 2 \textstyle\sum_{i=1}^{k-1} \eta_i B \eta_i^t
\end{aligned} \tag{2.1}$$

To show $B \ge 0$, denote $B = (b_{ij})$. Then $b_{ii} = \sum_{j=1}^n a'_{ij}$, $b_{ij} = -a'_{ij}$, $i \ne j$. Let λ be an eigenvalue of B. by Gersgorin's Theorem ([26] pp 226), we know $|\lambda - b_{ii}| \le$

$\sum_{j=1, j\neq i}^{n} |b_{ij}|$, i e , $|\lambda - \sum_{j=1}^{n} a'_{ij}| \leq \sum_{j=1, j\neq i}^{n} a'_{ij}$ Therefore, $\lambda \geq 0$, which means $B \geq 0$ We complete the proof □

Since discrete programming is $NP-$ hard, we should consider using a continuous critical function to approximate the discrete one Next we give the corresponding continuous critical problem which is obtained by relaxing the conditions in the discrete critical problem

Critical Problem 2 *Find the maximum of* $f(\eta_1, \ldots, \eta_{k-1})$ *with conditions* $\|\eta_i\|^2 = \frac{n}{2k}$ $\eta_i \in R^n$ *and* $< \eta_i, \eta_j >= 0$ *for* $i \neq j$ $i, j = 1, \ldots, k-1$

Since B is symmetric, the eigenvalues of B are real and the eigenvectors associated with distinct eigenvalues of B are orthogonal([26], pp 76) Therefore we can denote $\lambda_1(B) \geq \lambda_2(B) \ldots \geq \lambda_n(B)$ as eigenvalues of B and $e_1(B), e_2(B), \ldots, e_n(B)$ as their corresponding eigenvectors with $\|e_i(B)\|^2 = \frac{n}{2k}$ The following lemma shows the existence of the solution of Critical Problem 2

Lemma 7 *The solution of the* Critical Problem 2 *is* $\eta_i = e_i(B)$ $i = 1, \ldots, k-1$ *and* $f(\eta_1, \ldots, \eta_{k-1}) = \frac{n}{k}\sum_{l=1}^{k-1} \lambda_l(B)$

Proof By using the method of Lagrange Multipliers ([11], pp 348) Critical Problem 2 is equivalent to finding the maximum of the following function

$$F = f - \sum_{l=1}^{k-1} \mu_l \|\eta_l\|^2 - \sum_{1\leq i<j\leq k-1} \nu_{ij} < \eta_i, \eta_j >,$$

where μ_i and ν_{ij}, $1 \leq i, j \leq k-1$ are parameters to be defined

If we define $\frac{\partial F}{\partial \eta_i} = (\frac{\partial F}{\partial \xi_1^i}, \frac{\partial F}{\partial \xi_2^i}, \ldots, \frac{\partial F}{\partial \xi_n^i})^t$, then we have

$$\frac{\partial F}{\partial \eta_i} = 4B\eta_i^t - 2\mu_i\eta_i^t - 2\sum_{i \neq j} \nu_{ij}\eta_j^t = 0 \tag{1}$$

We derive $\frac{\partial <\eta_i, B\eta_i>}{\partial \eta_i} = 2B\eta_i^t$. The other parts can be obtained in the same way. In fact,

$$\begin{aligned} \frac{\partial <\eta_i, B\eta_i>}{\partial \xi_j^i} &= \frac{\partial \sum_{1<l,m<n} b_{lm}\xi_l^i\xi_m^i}{\partial \xi_j^i} \\ &= 2\sum_{m=1}^{n} b_{jm}\xi_m^i, \quad j = 1, 2, \ldots, n \end{aligned} \tag{2.2}$$

By multiplying $\frac{1}{2}\eta_i$ to equation (1), we get

$$2\eta_i B\eta_i^t - \mu_i\|\eta_i\|^2 - \sum_{i \neq j} \nu_{ij} < \eta_i, \eta_j >= 0$$

Since $< \eta_i, \eta_j >= 0$, $i \neq j$, so $2\eta_i B\eta_i^t - \mu_i\|\eta_i\|^2 = 0$, i.e., $2\sum_{i=1}^{k-1} \eta_i B\eta_i^t - \sum_{i=1}^{k-1} \mu_i\|\eta_i\|^2 = 0$, therefore $f = \frac{n}{2k}\sum_{i=1}^{k-1} \mu_i$

We notice that the maximum has nothing to do with ν_{ij}, so we can set $\nu_{ij} = 0$ in equation (1). Then

$$B\eta_i^t = \frac{1}{2}\mu_i\eta_i^t, \quad i = 1, \ldots, k-1$$

Hence $\frac{1}{2}\mu_i$ and η_i are eigenvalues and eigenvectors of B. Therefore, to obtain the maximum and also satisfy the given conditions, we choose $k-1$ largest eigenvalues and the corresponding eigenvectors. We have proven the lemma. □

Let $A = (a_{ij})$ be the adjacency matrix of G'. Then for $B = D - A$, $e_1(B), e_2(B), \ldots, e_{k-1}(B)$ are an optimal solution for Critical Problem 2 for the case $Q = G'$. Denote eigenvalues of A by $\lambda_n(A) \geq \lambda_{n-1}(A) \geq \ldots \geq \lambda_1(A)$, and the corresponding orthogonal eigenvectors $e_n(A), e_{n-1}(A), \ldots, e_1(A)$, $\|e_i(A)\|^2 = \frac{n}{2k}$, $i = 1, 2, \ldots, n$. The following lemma will allow us to apply Alon and Kahale's results [2].

Lemma 8 *Almost surely* $\lambda_i(B) = \lambda_i(D - A) = (1 + o(1))(k-1)d - \lambda_i(A)$, $d = \frac{np}{k}$, $i = 1, 2, \ldots, n$.

Proof From Lemma 1, we know that the expectation of d_{ii} is $(k-1)d$. By noticing the Rayleigh quotient $R(B) = R(D - A) = \frac{<x,(D-A)x>}{<x,x>} = \frac{<x,Dx>}{<x,x>} + \frac{<x,(-A)x>}{<x,x>} = \frac{<x,(1+o(1))(k-1)dx>}{<x,x>} + R(-A) = (1 + o(1))(k-1)d\frac{<x,x>}{<x,x>} + R(-A) = (1 + o(1))(k-1)d + R(-A)$, and applying the Courant-Fischer theorem, we obtain the lemma. □

Therefore, we can use $e_1(A), e_2(A), \ldots, e_{k-1}(A)$ as an optimal solution for Critical Problem 2 with a relatively small error.

Critical functions, which depend on the graph's adjacency matrix, have been used for the study of various graph properties [9, 21]. How to relax restrictions on the critical function so that a correctness proof is possible is the key issue to resolve when applying spectral theory to the study of graphs.

Chapter 3

Algorithm for coloring random k-colorable graphs

Using previous chapter's results, we present an algorithm for coloring random k-colorable graphs.

Without loss of generality, we can order all vertices in such way that vertex $v_{i+(j-1)\frac{n}{k}}$ is colored by $j-th$ color, here $i = 1, 2, \ldots, \frac{n}{k}$, $j = 1, \ldots, k$. Therefore, we can assign all vertices of V to k color points stated in Lemma 4, i.e. map $v_{i+(j-1)\frac{n}{k}}$ to u_j^k. In the other word, $v_1, v_2, \ldots, v_{\frac{n}{k}}, v_{1+\frac{n}{k}}, v_{2+\frac{n}{k}}, \ldots, v_{\frac{n}{k}+\frac{n}{k}}, \ldots, v_{1+(k-1)\frac{n}{k}}, v_{2+(k-1)\frac{n}{k}}, \ldots, v_n$ correspond to $u_1^k, u_1^k, \ldots, u_1^k, u_2^k, u_2^k, \ldots, u_2^k, \ldots, u_{k-1}^k, u_{k-1}^k, \ldots, u_{k-1}^k$. Denote $W_j = (w_1^j, \ldots, w_n^j)$ with $w_{i+(l-1)\frac{n}{k}}^j = r_j^{l,k}$, $i, l = 1, \ldots, k$, $j = 1, \ldots, k-1$. W_j consists of all jth coordinates of $u_1^k, u_1^k, \ldots, u_1^k, u_2^k, u_2^k, \ldots, u_2^k, \ldots, u_{k-1}^k, u_{k-1}^k, \ldots, u_{k-1}^k$. By noticing that each color class contains the same number of vertices, and $(W_i, W_j) =$

$k(U_i^k, U_j^k) = 0$, we know that $\{W_i\}_1^{k-1}$ is an orthogonal set in R^n Similarly, $\sum_{i=1}^n w_i^j = 0$

In the next chapter, we will show that there exists a matrix T such that

$$(W_1, W_2, \ldots, W_{k-1})^t = (e_1, e_2, \ldots, e_{k-1})^t T + (\eta_1, \ldots, \eta_{k-1})^t$$

where $\|\eta_i\|^2 = O(\frac{n}{d})$, $TT^t = I$, I is the identity matrix and $e_1, \ldots e_{k-1}$ are the smallest eigenvectors of the adjacency matrix A Since applying an orthogonal transformation to n given points does not change their relative locations a simple way to generate an approximation to a proper coloring can be given

If we denote $e_i = (y_1^i, \ldots, y_n^i)$ and $v_i = (y_i^1, \ldots, y_i^{k-1})$, and if we put all those points, which are corresponding to vertices $v_1, \ldots, v_n$ in R^{k-1}, then we can find a rotation on those points $u_1^k, \ldots, u_k^k$ such that for most vertices in V, each vertex is near to one of the color points $\{u_1^k, \ldots, u_k^k\}$ According to this property, all vertices except a small fraction can be colored properly

By using this result, we can propose the following modified algorithm based on Alon and Kahale's[2] The proof of its correctness appears in the following section

Phase 1 First calculate the distance between every two distinct points corresponding to $\{v_1, v_2, \ldots v_n\}$ Second find one point from $\{v_1, v_2, \ldots v_n\}$ such that there are $\frac{n}{k} \pm O(\frac{n}{d})$ points which have less than $\frac{1}{2}$ distance to this point Then color those vertices corresponding to those points by color 1. Delete all those colored vertices from V and repeat the above process on the remaining vertices for $k - 1$ times until all k colors have been used.

Phase 2 Let Phase 1 be the $0th$ iteration. In the ith iteration, $0 < i \leq q = [\log_2 n]$, construct the color classes V_l^i, $l = 1, \ldots, k$, as follows. For every vertex v of G, let $N(G)$ denote the set of all its neighbors in G. In the ith iteration, color v by the least popular color of its neighbors in the previous iteration. That is, put v in V_j^i if $|N(v) \bigcap V_j^{i-1}|$ is the minimum among the k quantities $|N(v) \bigcap V_l^{i-1}|$, $(l = 1, \ldots, k)$, where ties are broken arbitrarily.

Phase 3 This phase consists of two stages. First, repeatedly uncolor every vertex colored j that has less than $\frac{d}{2}$ neighbors (in G), colored l, for some $l \in \{1, \ldots, k\} - \{j\}$. Recall that $d = \frac{np}{k}$. Then if the graph induced on the set of uncolored vertices has a connected component of size larger than $\log_k n$, the algorithm fails. Otherwise, find a coloring of every component consistent with the rest of the graph using exhaustive search. If the algorithm can not find such a coloring, it fails.

Chapter 4

The Correctness Proof of the Algorithm

In this chapter, we will show the correctness of the algorithm which can be concluded as the following theorem.

Theorem 4 *If $p > \frac{c}{n}$, $c \gg 1$, then the algorithm produces a proper k-coloring of G with probability $1 - o(1)$.*

We will verify the algorithm in three steps. The approach used here is similar to Alon and Khale's [2].

§4.1 Initial Approximation Solution

Since the eigenvector set of A spans R^n and a proper coloring can be seen as $k-1$ vectors in R^n, then a proper coloring corresponds to linear combinations of the eigenvectors. We can see that it is still an $NP-$ hard problem to find these coloring vectors in $n-$dimensional space. The intuition is to find the proper coloring in a space with constant dimensions which depends on k instead of finding them in the $n-$dimensional space.

With a slight modification in the proof of Lemma 2.1[2] (just changing "3" to "k"), we have the following lemma which bounds eigenvalues of A.

Lemma 9 *Almost surely,*

(i) $\lambda_n(A) \geq (1 - 2^{-\Omega(d)})(k-1)d$

(ii) $\lambda_1(A) \leq \cdots \leq \lambda_{k-1}(A) \leq -(1 - 2^{-\Omega(d)})d$

(iii) $|\lambda_i| \leq O(\sqrt{d})$ *for all* $k \leq i \leq n-1$

Therefore we can express $\{W_j\}_1^{k-1}$ as a linear combination of $\{e_j\}_1^{k-1}$ with a small perturbation almost surely. We state this as follows.

Lemma 10 *Almost surely*

$$(W_1, W_2, \ldots, W_{k-1})^t = (e_1, e_2, \ldots, e_{k-1})^t R + (\eta_1, \ldots, \eta_{k-1})^t \qquad (4)$$

where $\|\eta_i\|^2 = O(\frac{n}{d})$, $RR^t = I + \Delta I$, $\|\Delta I\|^2 = O(\frac{1}{d})$, *I is the identity matrix*

Proof We show that $W_1^t = (e_1, e_2, \ldots, e_{k-1})^t R + \eta_1^t$ The remaining ones can be proven similarly Alon and Kahale [2] gave the outline of the proof We will complete the proof

We first claim that $\|(A + dI)W_1\|^2 = O(nd)$ In fact, it is sufficient to prove that the sum of squares of the coordinates of $(A + dI)W_1$ on the first color set is $O(nd)$ almost surely, as the sums on other color sets can be bounded similarly We first show that the expectation of the square of each coordinate of $(A + dI)W_1$ is $O(d)$ Denote $(A + dI)W_1 = (q_i)_1^n$ Then $q_i = dw_i^1 + \sum_{j=1, j\neq i}^{n} a_{ij} w_j^1$ For $i = 1$, we have $E(q_1) = dw_1^1 + E(\sum_{j=1, j\neq i}^{n} a_{ij} w_j^1) = dw_1^1 + \sum_{j=1+\frac{n}{k}}^{n} p w_j^1$ Since $w_1^1 = w_2^1 = \cdots = w_{\frac{n}{k}}^1$ so $dw_1^1 = \frac{np}{k} w_1^1 = p \sum_{j=1}^{\frac{n}{k}} w_j^1$ Hence, $E(q_1) = p \sum_{j=1}^{\frac{n}{k}} w_j^1 + \sum_{j=1+\frac{n}{k}}^{n} p w_j^1 = p \sum_{j=1}^{n} w_j^1 = 0$ In the same way, we can show $E(q_i) = 0$ for $i > 1$ Since $var(q_i) = O((k-1)d(1-p)) = O(d)$ and $var(q_i) = E(q_i - E(q_i))^2$, therefore $E(q_i)^2 = var(q_i) = O(d)$ Similarly the expectation of the fourth power of each coordinate of $(A + dI)W_1$ is $O(d^2)$ Hence, the variance of the square of each coordinate is $O(d^2)$ However the coordinates of $(A + dI)W_1$ on the first color set are independent random variables, and hence the variance of the sum of the squares of the first color set coordinates is equal to the sum of the variances which is $O(nd^2)$ If we denote the random variable $\|(A + dI)W_1\|^2$ by X, then $E(X) = O(nd)$ and $\sigma_X = O(\sqrt{n}d)$ From Chebyshev's Inequality, we have $P(|X - E(X)| > n^{\frac{1}{4}}\sigma_X) < \frac{1}{\sqrt{n}}$ Therefore, almost surely $|X - E(X)| \leq n^{\frac{1}{4}}\sigma_X$, i e , $|X - E(X)| = O(n^{\frac{3}{4}}d)$ Since $E(X) = O(nd)$, $X = O(nd)$, i e , $\|(A + dI)W_1\|^2 = O(nd)$ almost surely

Since $e_1, \quad , e_n$ are a basis for R^n, we have $W_1 = \sum_{i=1}^n c_i e_i$ Then $(A + dI)W_1 = \sum_{i=1}^n c_i(\lambda_i + d)e_i$ And by using Lemma 9, we have

$$\|(A + dI)W_1\|^2 = \sum_{i=1}^{n}(\lambda_i + d)^2 = \Omega(d^2)\sum_{i=k}^{n} c_i^2$$

Therefore $\|\eta_i\|^2 = \sum_{i=k}^n c_i^2 = O(\frac{n}{d})$ Now we show $RR^t = I + \Delta I$ $\|\Delta I\|^2 = o(\frac{1}{d})$

From Lemma 4 we know that $\|W_1\|^2 = \frac{n}{2k}$ In Chapter 2 we choose $\|e_i\|^2 = \frac{n}{2k}$ $i = 1, 2, \quad n$ In equation (4) both sides multiply their own transpositions, the left-hand side becomes $\frac{n}{2k}$, the right-hand side becomes $\frac{n}{2k}RR^t + \Delta M$ with $\|\Delta M\|^2 = O(\frac{n}{d})$ Combining the claim and the inequality, we prove the lemma □

Next we show that if a matrix is almost orthogonal, then it can be decomposed to an orthogonal matrix and a "smaller" matrix

Lemma 11 *If $RR^t = I + \Delta I$ with $\|\Delta I\|^2 = o(\frac{1}{d})$, then $R = T + \Delta T$ with $TT^t = I$ and $\|\Delta T\|^2 = o(\frac{1}{d})$*

Proof Denote R by $[r_1 \quad , r_{k-1}]$ From the given conditions, we have $\|r_i\|^2 = 1 + o(\frac{1}{d})$ and $| < r_i \ r_j > | = o(\frac{1}{d})$, $i \neq j$ If we construct $T = [t_1 \quad t_{k-1}]$ by the standard iterative procedure as follows

$$t_1 = r_1 \quad t_1 = \frac{1}{\|t_1\|}t_1$$

$$\tilde{t}_i = <t_1, r_i> t_1 + \cdots + <t_{i-1}, r_i> t_{i-1} - r_i,$$
$$t_i = \frac{1}{\|\tilde{t}_i\|}\tilde{t}_i, \; i = 2, \ldots, k-1,$$

then $TT^t = I$. Next we prove $R = T + \Delta T$ with $\|\Delta T\|^2 = o(\frac{1}{d})$. We first claim that $t_i = r_i + \Delta t_i$ with $\|\Delta t_i\|^2 = o(\frac{1}{d})$, $i = 1, 2, \ldots, k-1$. We prove this claim by induction. For $i = 1$, since $\|r_1\|^2 = 1 + o(\frac{1}{d})$, $\frac{1}{\|r_1\|} = 1 + o(\frac{1}{d})$. Therefore, $t_1 = \frac{1}{\|r_1\|}r_1 = r_1 + o(\frac{1}{d})r_1$. Suppose that the claim is true for $i = m-1$. Then for $i = m$, from the induction assumption $t_i = r_i + \Delta t_i$ with $\|\Delta t_i\|^2 = o(\frac{1}{d})$, $i = 1, 2, \ldots, m-1$, we have

$$\begin{aligned} &|<t_i, r_m>| = |<r_i, r_m> + <\Delta t_i, r_m>| \\ &\le |<r_i, r_m>| + |<\Delta t_i, r_m>| \\ &\le o(\tfrac{1}{d}) + \|\Delta t_i\|\|r_m\| = o(\tfrac{1}{d}), \end{aligned} \tag{4.1}$$

and

$$\begin{aligned} &\|\tilde{t}_m\|^2 = \|\textstyle\sum_{l=1}^{m-1} <t_l, r_m> t_l - r_m\|^2 \\ &\le \|\textstyle\sum_{l=1}^{m-1} <t_l, r_m> t_l\|^2 + \|r_m\|^2 \\ &\le \textstyle\sum_{l=1}^{m-1} |<t_l, r_m>| \|t_l\|^2 + \|r_m\|^2 \\ &= 1 + o(\tfrac{1}{d}) \end{aligned} \tag{4.2}$$

Therefore, from $\tilde{t}_m = \sum_{l=1}^{m-1} <t_l, r_m> t_l - r_m$ and above inequality we have $t_m = r_m + \Delta t_m$. We have proven the lemma. □

From the above two lemmas, we deduce that

Lemma 12 *Almost surely, there exists an orthogonal matrix T such that*

$$(W_1, W_2, \ldots, W_{k-1})^t = (e_1, e_2, \ldots, e_{k-1})^t T + (\eta_1, \ldots, \eta_{k-1})^t$$

where $\|\eta_i\|^2 = O(\frac{n}{d})$, $TT^t = I$, *I is the identity matrix*

Therefore $(W_1, W_2, \ldots, W_{k-1})$ can be obtained from an orthogonal rotation of $(e_1, \ldots, e_{k-1})^t$ with a small perturbation $O(\frac{n}{d})$.

§4.2 The Iterative Procedure

In this section we will show that after the second phase, almost surely all vertices in a subset of V are colored properly. As we have shown in the last section, after the first phase all vertices except $O(\frac{n}{d})$ are colored properly. Then we can show that after each iteration, the size of vertex set which receives wrong colors will be reduced at least by half. This can be seen from the fact that almost every vertex has a relatively large degree ($(k-1)d >> 1$), hence if we have a relatively small set wrongly colored, its vertices will have fewer neighbors.

Construct H as follows. First, set H_0 to be the subset of V with vertices having at most $1.01d$ neighbors in V in each color class. Then set H_i to be the subset of H_{i-1} by deleting any vertex in H_{i-1} having less than $0.99d$ neighbors in H_{i-1} in some color class (other than its own.) $i = 1, 2, \ldots$. The final result is denoted by H.

By using the Chernoff Bound Theorem, we can deduce the following lemma.

Lemma 13 *For $\epsilon = 0.001$, there exists a constant $\gamma > 0$ such that the follows hold almost surely*

1. *For any two distinct color classes V_1 and V_2, and any subset X of V_1 and any subset Y of V_2, if $|X| = 2^{-\gamma d}\frac{n}{k}$ and $|Y| \leq k|X|$, then $|e(X, V_2 - Y) - d|X|| \leq \epsilon d|X|$*

2. *If J is the set of vertices having more than $1.01d$ neighbors in G in some color class, then $|J| \leq 2^{-\gamma d}\frac{n}{k}$*

Proof We first prove (1). Since

$$\begin{aligned}\mu &= E(e(X, V_2 - Y)) = |V_2 - Y||X|p \\ &> [\tfrac{n}{k} - k|X|]|X|p \\ &= [\tfrac{1}{k} - 2^{-\gamma d}n]d2^{-\gamma d}n,\end{aligned} \tag{4.3}$$

if we set $\gamma > \frac{\log_2 k}{c}$, similar to the proof of Lemma 2, we have $\lim_{n\to\infty} \mu = \infty$. Furthermore,

$$\begin{aligned}&P(e(X, V_2 - Y)) > (1+\epsilon)d|X|) = P(e(X, V_2 - Y)) > (1+\epsilon)\tfrac{n}{|V_2 - Y|k}\mu) \\ &\leq P(e(X, V_2 - Y)) > (1+\epsilon)\mu),\end{aligned} \tag{4.4}$$

therefore, $\lim_{n\to\infty} P(e(X, V_2 - Y)) > (1+\epsilon)d|X|) = 0$, i.e., $e(X, V_2 - Y) - d|X| \leq \epsilon d|X|$ almost surely. By using Theorem 4.2 in [28], pp.87, we have $P(e(X, V_2 - Y) < (1-\epsilon)d|X|) < e^{-\mu\epsilon^2/2}$. Therefore, $\lim_{n\to\infty} P(e(X, V_2 - Y) < (1-\epsilon)d|X|) = 0$, i.e., $e(X, V_2 - Y) - d|X| \geq -\epsilon d|X|$ almost surely. Thus we have proven (1).

Now we prove (2). Suppose there exists a set J having more than $1.01d$ neighbors in G in some color class, say V_2, with $|J| > 2^{-\gamma d}\frac{n}{k}$. Set $X = J$ and $Y = 0$. Then we have $|e(X, V_2 - Y) - d|X|| > 0.1d|X|$, which contradicts to (1). Therefore, (2) is true. □

Replacing 3 by k in the proof of Lemma 3.5 [2], we have

Lemma 14 *[2] Almost surely, H has at least $(1 - 2^{-\Omega(d)})\frac{n}{k}$ vertices in every color class.*

Lemma 15 *Almost surely, there are no two subsets of vertices U and W of V such that $|U| \leq 0.001\frac{n}{k}$, $|W| = |U|/2$, and every vertex v of W has at least $d/4$ neighbors in U.*

Proof Note that if there are such two (not necessarily disjoint) subsets U and W, then the number of edges joining vertices of U and W is at least $d|U|/8$ with $|U| \geq \frac{d}{4}$. Therefore, the probability that there exist such two subsets is at most

$$
\begin{aligned}
&\left(\sum_{i=0}^{|U|} \binom{|U|}{\frac{d}{4}+i} p^{\frac{d}{4}+i}\right)^{|W|} \\
&\leq \left(\sum_{i=0}^{|U|} \left(\frac{|U|e}{\frac{d}{4}+i}\right)^{\frac{d}{4}+i} p^{\frac{d}{4}+i}\right)^{\frac{|U|}{2}} \\
&= \left(\left(\sum_{i=0}^{|U|} \left(\frac{|U|e}{\frac{d}{4}+i}\right)^{\frac{d}{4}+i} \left(\frac{dk}{n}\right)^{\frac{d}{4}+i}\right)^{\frac{|U|}{2}}\right. \\
&\leq \left(\sum_{i=0}^{|U|} \left(\frac{4ek|U|}{n}\right)^{\frac{d}{4}+i}\right)^{\frac{|U|}{2}} \qquad (4.5) \\
&\leq \left(\left(\frac{4ek|U|}{n}\right)^{\frac{d}{4}}\right)^{\frac{|U|}{2}} \\
&\leq \left(\frac{2ekd}{n}\right)^{\frac{d^2}{8}} \\
&= O(1/n^{\Omega(d)})
\end{aligned}
$$

The lemma follows immediately □

From this lemma, we can see that if the set of wrongly colored vertices is relatively small, we can reduce its number at least by half through each iteration. This is summarized as the following lemma.

Lemma 16 *[2] Almost surely, by the end of the second phase of the algorithm, all vertices in H are properly colored.*

§4.3 Uncoloring and Exhaustive Search

After the iterative phase, we know that all vertices in H, which is pruned so that the edge number in each color class (except its own class) is about d, receive proper colors almost surely. In this section, we will show that a largest connected component of the graph induced by $V - H$ has at most $O(\log_k n)$ vertices. Hence, we can use exhaustive method to color the remaining vertices which have been uncolored in the first step of the third phase.

The next lemma shows that the vertex colors in the graph H remain unchanged after the uncoloring process almost surely.

Lemma 17 *[2] Almost surely, by the end of the uncoloring procedure in Phase 3 of the algorithm, all vertices of H remain colored, and all colored vertices are properly* *colored, i.e. any vertex colored i belongs to W_i. (We assume, of course, that the numbering of the colors is chosen appropriately)*

Let T be a fixed tree on $\log_k n$ vertices of V. Let I be the subset of all vertices $v \in V(T)$ whose degree in T is at most $k+1$. Let H' be the subset of V obtained by the following procedure, which is similar to the way generating H. First, set H'_0 to be the subset of V with vertices having at most $1.01d - (k+1)$ neighbors in V in each color class, then delete from it all vertices of $V(T) - I$. Then set H'_i to be the subset of H'_{i-1} by deleting any vertex in H'_{i-1} having less than $0.99d$ neighbors in H'_{i-1} in some color class, $i = 1, 2, \ldots$. Denote H' the final result.

Lemma 18 *[2] Let F be any subset of edges with endpoints in V, then $H'_i \subseteq H_i$, $i = 0, 1, 2, \ldots$. Hence, $H' \subseteq H$.*

Lemma 19 *[2]*

$$\begin{aligned} &Pr[T \text{ is a subgraph of } G \text{ and } V(T) \bigcap H = 0] \\ &\leq Pr[T \text{ is a subgraph of } G] \, Pr[I \bigcap H' = 0] \end{aligned} \tag{4.6}$$

Lemma 20 *Almost surely the largest connected component of the graph induced on $V - H$ has at most $\log_k n$ vertices.*

Proof Let T be a fixed tree on $\log_k n$ vertices of V Similarly to Lemma 14, it can be shown that H' missed at most $2^{-\Omega(d)}n$ vertices in each color class Therefore the probability $Pr[I \bigcap H' = 0]$ is at most $2^{-\Omega(d|I|)}$ Since $|I| \geq \frac{|V(T)|}{2}$ and the probability $Pr[\text{T is a subgraph of G}] = (\frac{d}{n})^{|V(T)|-1}$, thus the probability for T is a subgraph of size $\log_k n$ which is a connected component of the induced subgraph of G on $V - H$ is at most $2^{-\Omega(d\log_k \frac{n}{2})}(\frac{d}{n})^{\log_k n-1}\binom{n}{\log_k n}(\log_k n)^{\log_k n-2}$ Therefore from Lemma 14 we have

$$
\begin{aligned}
& Pr[\text{T is a subgraph of G and } V(T) \bigcap H = 0] \\
& \leq Pr[\text{T is a subgraph of G}]\ Pr[I \bigcap H' = 0] \\
& \leq 2^{-\Omega(d\log_k \frac{n}{2})}(\tfrac{d}{n})^{\log_k n-1}\binom{n}{\log_k n}(\log_k n)^{\log_k n-2}2^{-\Omega(d\log_k n)}n \\
& = O(n^{-\Omega(d)})
\end{aligned}
\tag{4.7}
$$

Hence we complete the proof

Chapter 5

Conclusion

The heuristic graph coloring approach we proposed in Chapter 2, which is equivalent to the initial approximate solution given by Alon and Kahale [2] for the 3−colorable graph case, has a similar flavor to techniques used in the study of the eigenvalues of the Laplacian of graphs [9]. The correctness proof of the algorithm is based on evaluating all eigenvalues of the adjacency matrix given by Alon and Kahale. We have shown that the initial approximate solution can be obtained through a critical function. The existence of an approximation algorithm based on a critical problem for coloring arbitrary graphs is a question that deserves further investigation.

Bibliography

[1] N Alon, M Blum A Fiat, S Kannan, M Naor and R Ostrovsky Matching nuts and bolts (preprint)

[2] N Alon and N Kahale, A spectral technique for coloring random 3-colorable graphs In Proceedings of the Twenty-Sixth ACM Symposium on Theory of Computing, pp 346-353, 1994

[3] N Alon and J H Spencer, The Probabilistic Method, Willey Interscience, 1992

[4] R Bellman, Introduction to Matrix Analysis, McGraw-Hill Book Company Inc , 1960

[5] A Blum New approximation algorithms for graph coloring Journal of ACM, 41 470-516, 1994

[6] B Bollobas Random Graphs Academic Press, 1985

[7] B Bollobas, The chromatic number of random graphs, Combinatorica, 8(1) 49-55, 1988

[8] A Borodin and J E Hopcroft, Routing, merging, and sorting on parallel models of computation, Journal of Computer and System Sciences, 30, 130-145, 1985

[9] F R K Chung and S T Yao, A near optimal algorithm for edge separators, Proc 26^{th} STOC ACM Press 1-8, 1994

[10] T H Cormen C E Leiserson and R L Rivest, Introduction to Algorithms, MIT Press, 1990

[11] L J Corwin and R H Szczarba, Calculus in Vector Spaces, Marcel Dekker Inc 1979

[12] J Culberson, Iterated greedy graph coloring and the difficulty landscape, Technical Report TR 92-07, University of Alberta, Department of Computer Science, Canada

[13] J Ellis and P Lepolesa, A Las Vegas coloring algorithm, The Computer Journal 32(5) 474-476 1989

[14] J Friedman J Kahn and E Szemeredi, On the second eigenvalue in random regular graphs Proc 21^{st} ACM STOC ACM Press (1989) 587 598

[15] M Garey and D Johnson, Computers and Intractability A Guide to the Theory of NP-Completeness, W H Freeman and Company San Francisco, 1979

[16] M Garey, D Johnson and L Stockmeyer, Some simplified NP-complete graph problems, Theoret Comput Sci 1(1976), 237-267

[17] G Grimmett and C McDiarmid, On coloring random graphs, Mathematical Proceedings of the Cambridge Philosophical Society, 77 313-324 1975

[18] M M Halldorsson, A still better performance guarantee for approximate graph coloring Information Processing Letters 45 19-23 1993

[19] R A Horn and C R Johnosn Matrix Analysis, Cambridge Univ Press, New York (1985)

[20] D S Johnson Worst case behavior of graph coloring algorithms Proc 5th South-eastern Conf on Combinatorics, Graph Theory and Computing, Congressus Numerantium X, 513-527, 1974

[21] D Karger R Motwani and M Sudan Approximate graph coloring by semidefinite programming (to appear)

[22] R M Karp Reducibility among combinatorial problems in "Complexity of Computer Computations"(R E Miller and J W Thatcher Eds) Plenum, New York 1972

[23] V. King, A lower bound for the recognition of digraph properties, Combinatorica, 10(1), 53-59, 1990.

[24] V. King, An $\Omega(n^{\frac{5}{4}})$ lower bound on the randomized complexity of graph properties, Combinatorica, 11(1), 23-32, 1991.

[25] Donald E. Knuth, Foundamental Algorithms, The Art of Computer Programming Vol. 1. Addison Wesley, Reading, MA, 1971.

[26] P. Lancaster. The Theory of Matrices, Academic Press Inc., 1969.

[27] L. Lovasz. On the Shannon capacity of a graph. IEEE Transactions on Information Theory, IT-25 1-7, 1979.

[28] R. Motwani and P. Raghavan. Randomized Algorithms. Cambridge, New York: Cambridge University Press, 1995.

[29] L. Stockmeyer, Planar 3-colorability is NP-complete, SIGACT News, 19-25, 1973.

[30] J. Turner. Almost all k-colorable graphs are easy to color, Journal of Algorithms, 9 63-82, 1988.

[31] A. Wigderson. Improving the performance guarantee for approximate graph coloring. Journal of the ACM, 30 729-735, 1983.

[32] S. Mahajan and H. Ramesh, Derandomizing semidefinite programming based approximation algorithms, 36th IEEE Symposium on Foundations of Computer Science, pp. 162-169, 1995.

VITA

Surname: Yan

Given Names: Peter Pengxiang

Place of Birth: Xiamen, China

Date of Birth: 1961

Educational Institutions Attended

University of Victoria	1993 to 1996
Fuzhou University	1978 to 1982

Degrees Awarded

B Sc	Fuzhou University	1982

Title of Thesis

Coloring Random $K-$ Colorable Graphs

Author [redacted]

Peter Pengxiang Yan

August 20, 1996