Randomized Distributed Edge Coloring via an Extension of the Chernoff-Hoeffding Bounds*

Alessandro Panconesi
Aravind Srinivasan

TR 93-1357
June 1993

Department of Computer Science
Cornell University
Ithaca, NY 14853-7501

*A preliminary version of this work appears as "Fast Randomized Algorithms for Distributed Edge Coloring", in the *ACM Symposium on Principles of Distributed Computing,* pages 251-262, August 1992. This research was supported in part by NSF PYI award CCR-89-96272 with matching support from UPS and Sun Microsystems.

Randomized Distrubuted Edge Coloring via an Extension of the Chernoff-Hoeffding Bounds *

Alessandro Panconesi & Aravind Srinivasan
Department of Computer Science, Cornell University
Ithaca, NY 14853
E-mail: {ap, srin}@cs.cornell.edu

June 9, 1993

Abstract

Certain types of routing, scheduling and resource allocation problems in a distributed setting can be modeled as edge coloring problems. We present fast and simple randomized algorithms for edge coloring a graph, in the synchronous distributed point-to-point model of computation. Our algorithms compute an edge-coloring of a graph G with n nodes and maximum degree Δ with at most $(1.6+\epsilon)\Delta+\log^{2+\delta} n$ colors with high probability (arbitrarily close to 1), for any fixed $\epsilon, \delta > 0$.

To analyze the performance of our algorithms, we introduce an extension of the *Chernoff-Hoeffding bounds*, which are fundamental tools that are used very frequently in estimating tail probabilities. However, they assume stochastic independence among certain random variables, which may not always hold. Our results extend the Chernoff-Hoeffding bounds to certain types of random variables which are not stochastically independent. We believe that these results are of independent interest, and merit further study.

*A preliminary version of this work appears as "Fast Randomized Algorithms for Distributed Edge Coloring", in the *ACM Symposium on Principles of Distributed Computing*, pages 251–262, August 1992. This research was supported in part by NSF PYI award CCR–89–96272 with matching support from UPS and Sun Microsystems.

1 Introduction

An important limitation for a distributed network without global memory in computing any function is that for an efficient algorithm, each processor can communicate only with those processors that are within a small radius around it. Models of parallel computation like the PRAM abstract this problem of locality away by assuming the existence of a global shared memory with fast concurrent access. We are interested in studying how fast individual processors can compute their portion of the output in a message–passing distributed system, with such "local" information alone. The model we study is the synchronous distributed point–to–point model, in which the processors are arranged as the vertices of an n–vertex graph $G = (V, E)$, and where all communication is via the edges of G alone.

The edge coloring problem can be used to model certain types of jobshop scheduling, packet routing, and resource allocation problems in a distributed setting. For example, the problem of scheduling I/O operations in some parallel architectures can be modeled as follows [6]. We are given a set of processes $\mathcal{P}$ and a set of resources $\mathcal{R}$ such that each process $p \in \mathcal{P}$ needs a subset $f(p) \subseteq \mathcal{R}$ of the resources where: *(i)* each process p needs every resource in $f(p)$ for a unit of time each, and *(ii)* p can use the resources in $f(p)$ in any order. From this, we can construct a bipartite graph $G_{\mathcal{P},\mathcal{R}} = (\mathcal{P}, \mathcal{R}, E_{\mathcal{P},\mathcal{R}})$ where $E_{\mathcal{P},\mathcal{R}} = \{(p, r) | \ p \in \mathcal{P} \wedge r \in f(p)\}$. An edge coloring of $G_{\mathcal{P},\mathcal{R}}$ with c colors yields a schedule for the processes to use the resources within c time units. Optimal colorings correspond to optimal schedules.

Edge coloring can also be used in distributed models in situations where broadcasts are infeasible or undesirable: an edge coloring of the network results in a schedule for each processor to communicate with at most one neighbor at every step (at time step i, processors communicate via the edges colored i only), and using a "small" number of colors reduces the wastage of time in this schedule.

Note that Δ colors are necessary to edge color a graph with maximum degree Δ. Vizing showed that it is always possible to edge color a graph with $\Delta + 1$ colors and gave a polynomial time algorithm [3]. Efforts to parallelize Vizing's theorem have failed so far. The best known result is an RNC algorithm of Karloff & Shmoys that uses $\Delta + \Delta^{0.5+\epsilon}$ colors. The algorithm can be derandomized in NC by standard techniques by Berger & Rompel, and Motwani, Naor & Naor [2, 10]. In the distributed model, the best known edge coloring algorithm is to apply a vertex coloring algorithm to the line graph of G. There are fast (polylogarithmic) randomized vertex coloring algorithms that use $(\Delta + 1)$ and Δ colors, which translate to $(2\Delta - 1)$– and $(2\Delta - 2)$– edge coloring algorithms respectively [9, 12]. There are no known polylogarithmic deterministic algorithms in the distributed setting for $(2\Delta - 1)$–edge coloring [1, 12]. Moreover, distributed Δ–edge coloring requires $\Omega(diameter(G))$ time, even with randomization[12].

We present fast and simple randomized algorithms to edge color G with at most $(1.6+\epsilon)\Delta + 0.4\log^{2+\delta} n$ colors with high probability for any fixed $\epsilon, \delta > 0$, where Δ is the maximum degree of the vertices of G. At the heart of our analysis is an extension of the Chernoff–Hoeffding bounds, which are key tools in bounding the tail probabilities of certain random variables [4, 5, 13].

The edge coloring algorithm is based on a very simple randomized procedure to color bipartite graphs. The algorithm can be explained in a few lines. Given a bipartite graph $G = (A, B, E)$

with maximum degree Δ, each vertex in B picks distinct colors from $\{1, 2, \ldots, \Delta\}$ at random for its edges without replacement, *i.e.*, edges incident to the same vertex get different colors. Then, each vertex $v \in A$ checks, for each color α, if more than one of its incident edges has color α and if so, chooses one of them at random as the *winner*, and all the other edges of color α which are incident to v are decolored. The key claim is that for every vertex, the number of decolored edges incident to it is at most $\Delta(1+\epsilon)/e$ with high probability, for any fixed $\epsilon > 0$ (e is the Euler number). Assuming that this holds, we can repeat the above iteration with a set of $\Delta(1+\epsilon)/e$ *fresh* colors, and so on. In spite of its simplicity the algorithm requires an interesting probabilistic analysis based on our extension of the Chernoff-Hoeffding bounds to a certain case of dependence among the random variables that we call λ–correlation. We believe that these results have the potential for further applications, and need further study.

2 Notation

A *message–passing distributed network* is an undirected graph $G = (V, E)$ where vertices, or nodes, correspond to processors and edges correspond to bi–directional communication links. Each node has its unique ID. The network is *synchronous*, *i.e.*, computation takes place in a sequence of *rounds*; in each round, each node reads messages sent to it by its neighbors in the graph, does any amount of local computation, and sends messages back to all of its neighbors. The time complexity of a distributed algorithm, or *protocol*, is given by the number of rounds needed to compute a given function

Notice that in this model the cost of sending a message from one vertex to another is proportional to the shortest path between the two vertices, There is no shared memory and processors can communicate only by sending messages through the network. Hence, if we want a protocol to run for t rounds, then each vertex can communicate only with vertices at distance at most t from it. This is not so in the PRAM model, where the shared memory allows any two processors to communicate in one unit of time

Given an undirected graph $G = (V, E)$ we denote by Δ its maximum degree, *i.e.*, the maximum number of edges incident with any node; by d_u we denote the degree of vertex u, and by $\delta(u)$ we denote the set of edges incident with vertex u.

Given a positive integer n, $[n]$ denotes the set $\{1, 2, \ldots, n\}$. Given a matrix M with c columns and r rows where $c \leq r$, the generic entry of M is denoted as M_{ik} where i denotes the *column* number and k the *row* number. (This is the reverse of the usual definition but it is more natural when we define the permanent of M.) The permanent of a (possibly non–square) matrix M is defined as the natural extension of the permanent of square matrices. Let $\mathcal{P} = \{\pi \mid \pi : [c] \to [r], \pi \text{ is one–to–one}\}$. Then,

$$perm(M) = \sum_{\pi \in \mathcal{P}} M_{1,\pi(1)}\, M_{2,\pi(2)}\, \cdots\, M_{c,\pi(c)}.$$

An event $\mathcal{A}$ is said to happen *with high probability* (w.h.p.) if $Pr(\mathcal{A}) \geq 1 - 1/f(n)$ for some superpolynomial function $f(n)$ (*i.e.*, $n^c = o(f(n))$ for all $c > 0$, and the failure probability of $\mathcal{A}$ is bounded by $1/f(n)$).

3 The Chernoff-Hoeffding Bounds Extension

In this section we introduce our extension of the Chernoff-Hoeffding bounds which are important tools used in estimating the tail probabilities of random variables. Given n *independent* random variables $X_1, X_2, \ldots, X_n$, these bounds are used in deriving an upper bound on the upper tail probability $Pr(X \geq (1+\epsilon)\mu)$, where $X = \sum_{i=1}^{n} X_i$, $\mu = E[X]$, and $\epsilon > 0$. We extend these bounds to a certain case of dependency among the X_i's, which we call λ–correlation. Since the general definition of λ–correlation is involved, we first introduce this notion in the case where the X_i's are 0-1 random variables and prove our extension of the Chernoff-Hoeffding bounds. This special case is simpler to handle and more intuitive but contains all the essential difficulties of the general case. After the special case is dealt with, we will take care of the more general case.

Before giving our extension of the Chernoff-Hoeffding bounds let us review Chernoff's classical approach to upper bound the upper tail probability of a random variable X, which is the sum of *independent* binary random variables $X_1, X_2, \ldots, X_n$ [4]. Chernoff's basic idea is to use Markov's inequality on the random variable e^{tX} for an arbitrary, but fixed, $t > 0$, and minimize with respect to t. That is, to use the fact that

$$\begin{aligned} Pr(X > (1+\epsilon)\mu) &= Pr(e^{tX} > e^{t(1+\epsilon)\mu}) \\ &\leq \frac{E[e^{tX}]}{e^{t(1+\epsilon)\mu}}, \end{aligned}$$

and minimize the last ratio for $t > 0$ [4]. This is achieved by finding a good upper bound for the numerator $E[e^{tX}]$ by using the fact that X is the sum of *independent* random variables. It is standard (see, *e.g.*, Raghavan[13]) to use this for showing that in this case,

$$\min_{t>0} \frac{E[e^{tX}]}{e^{t(1+\epsilon)\mu}} \leq F(\mu,\epsilon) \doteq \left[\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right]^{\mu}. \tag{1}$$

Hoeffding [5] considers a more general case where X is the sum of n independent and *bounded* random variables $X_i \in [a_i, b_i]$, and uses the above approach to show that if $E[X] = \mu$, then for $\epsilon > 0$,

$$\min_{t>0} \frac{E[e^{tX}]}{e^{t(1+\epsilon)\mu}} \leq G(\mu,\epsilon,\vec{a},\vec{b}) \doteq \exp\left[-\frac{2\,\mu^2\epsilon^2}{\sum_{i\in[n]}(b_i-a_i)^2}\right]. \tag{2}$$

Equations 1 and 2 will be used in our proofs. Henceforth, we refer to these bounds of Chernoff and Hoeffding as the CH-bounds. If ϵ is a fixed positive quantity no greater than 1 (which will be true in all our applications), then $F(\mu,\epsilon) \leq e^{-\epsilon^2\mu/3}$. Hence, if $\mu = \Omega(\log^{1+\delta} n)$, then $F(\mu,\epsilon)$ is the inverse of a superpolynomial function of n, for any fixed $\delta > 0$ (similar considerations apply to $G(\mu,\epsilon,\vec{a},\vec{b})$). This fact makes the CH-bounds a powerful tool for deriving strong performance guarantees for randomized algorithms and will be used repeatedly in this paper.

3.1 0-1 Random Variables

We now present our extension of the CH-bounds to the case of 0-1 λ–correlated random variables. We begin by defining λ–correlation[1] for 0-1 random variables.

Definition 1 *Let $X_1, X_2, \ldots, X_n$ be 0-1 random variables. The X_i's are* λ–correlated *if, for all nonempty $I \subseteq [n]$,*

$$Pr\left(\bigwedge_{i \in I} X_i = 1\right) \leq \lambda \prod_{i \in I} Pr(X_i = 1).$$

Before proving the extension of the CH-bounds, let us develop some intuition by giving an example of λ–correlation. Suppose we have Δ balls that are thrown uniformly and independently at random into Δ bins. Let B_i be an indicator random variable that is 1 if bin i is empty and 0 otherwise. These variables are 1–correlated. To see this, consider a subset $J \subseteq [\Delta]$ of bins and suppose that all these bins are empty. Then, given this information, the probability that some other bin, say i, remains empty decreases. That is, for all $J \subseteq [\Delta]$,

$$\begin{aligned} Pr(B_i = 1 \mid \bigwedge_{j \in J} B_j = 1) &= \left(1 - \frac{1}{\Delta - |J|}\right)^{\Delta} \\ &\leq \left(1 - \frac{1}{\Delta}\right)^{\Delta} \\ &= Pr(B_i = 1) \end{aligned}$$

by straightforward induction, this implies that for all nonempty $I \subseteq [\Delta]$,

$$Pr\left(\bigwedge_{i \in I} B_i = 1\right) \leq \prod_{i \in I} Pr(B_i = 1).$$

The CH-bounds say that if X is the sum of n *independent* 0-1 random variables then $Pr(X > (1+\epsilon)\mu) \leq F(\mu, \epsilon)$. The next theorem shows that if the X_i's are λ–correlated then $Pr(X > (1+\epsilon)\mu) \leq \lambda\ F(\mu, \epsilon)$. In the statement of the following theorem, μ is an upper bound on $E[X]$; this is sufficient because we are upper bounding the upper tail probability of X.

Theorem 1 *Let X be the sum of λ-correlated 0-1 random variables, and let $E[X] \leq \mu$. Then,*

$$Pr(X > (1+\epsilon)\mu) \leq \lambda\ F(\mu, \epsilon).$$

PROOF. As in the classical proof, we start by introducing a positive parameter t and by applying Markov's inequality to the variable e^{tX},

$$\begin{aligned} Pr(X > (1+\epsilon)\mu) &= Pr(e^{tX} > e^{t(1+\epsilon)\mu}) \\ &\leq \frac{E[e^{tX}]}{e^{t(1+\epsilon)\mu}}. \end{aligned}$$

[1]This was defined as "self–weakening with parameter λ" in [11].

To find a good upper bound on the numerator (we cannot use independence of the X_i's) we wright down its Taylor expansion, and use the fact that e^{tX} is bounded by e^{tn} to apply linearity of expectation to the infinite series:

$$\begin{aligned} E[e^{tX}] &= E\left[\sum_{k=0}^{\infty} \frac{t^k X^k}{k!}\right] \\ &= \sum_{k=0}^{\infty} \frac{t^k E[X^k]}{k!}. \end{aligned}$$

Focus on the generic term $E[X^k]$ of this series. From the fact that $X = \sum X_i$ and that the X_i's are 0-1 random variables it follows that $E[X^k]$ is a linear combination of the form

$$E[X^k] = \sum_{I \subseteq [n]} \alpha_I \, Pr\left(\bigwedge_{i \in I} X_i = 1\right)$$

for $I \neq \emptyset$, and where all α_I's are non-negative (this can be verified by unfolding X^k and by observing that $X_i^c = X_i$ for all i and all $c \neq 0$). In order to upper bound the term $E[X^k]$ we introduce n twin 0-1 random variables $\hat{X}_i$ which have the same distribution as the X_i's except that they are independent. From the hypothesis of λ–correlation and from the fact the twin variables are independent it follows that

$$\begin{aligned} Pr\left(\bigwedge_{i \in I} X_i = 1\right) &\leq \lambda \prod_{i \in I} Pr(X_i = 1) \\ &= \lambda \prod_{i \in I} Pr(\hat{X}_i = 1) \\ &= \lambda \, Pr\left(\bigwedge_{i \in I} \hat{X}_i = 1\right). \end{aligned}$$

Hence,

$$E[X^k] \leq \lambda \, E[\hat{X}^k].$$

By upper bounding each term of the series expansion of $E[e^{tX}]$ we get

$$E[e^{tX}] \leq \lambda \, E[e^{t\hat{X}}].$$

Notice now that $\hat{X}$ is the sum of n *independent* random variables and $E[\hat{X}] = E[X] \leq \mu$. Hence, by equation 1

$$\begin{aligned} Pr(X > (1+\epsilon)\mu) &\leq \frac{E[e^{tX}]}{e^{t(1+\epsilon)\mu}} \\ &\leq \lambda \frac{E[e^{t\hat{X}}]}{e^{t(1+\epsilon)\mu}} \\ &\leq \lambda \, F(\mu, \epsilon). \end{aligned}$$

□

To see an application of this theorem, let us go back to the balls-and-bins experiment, and suppose we are interested in estimating the number of empty bins after the balls are thrown; this number is given by the random variable $B = \sum_i B_i$. By linearity of expectation and by the fact that the balls are thrown at random independently of each other,

$$\begin{aligned} E[B] &= \sum_{i \in [\Delta]} E[B_i] \\ &= \sum_{i \in [\Delta]} \left(1 - \frac{1}{\Delta}\right)^{\Delta} \\ &\leq \frac{\Delta}{e}. \end{aligned}$$

We already saw that the B_i's are 1–correlated. Hence, by Theorem 1,

$$Pr(B > (1+\epsilon)\Delta/e) \leq F(\Delta/e, \epsilon).$$

Remark. Jain has proved the following lemma [14]: *Let $a_1, a_2, \ldots, a_n$ be n random trials (not necessarily independent) such that the probability that trial a_i 'succeeds' is bounded above by p regardless of the outcomes of the other trials. Then if X is the random variable that represents the number of 'successes' in these n trials, and Y is a binomial variable with parameters (n,p), then: $Pr[X \geq k] \leq Pr[Y \geq k]$, $0 \leq k \leq n$.*

The assumptions of Jain's lemma are strictly stronger than those of 1–correlation. For instance, in the balls and bins example,

$$Pr(B_\Delta = 1 | \bigwedge_{i \in [\Delta - 1]} B_1 = 0) = \frac{\Delta - 1}{\Delta + 1},$$

which, for $\Delta \geq 3$, is greater than $Pr(B_\Delta = 1)$ ($\approx 1/e$).

3.2 The General Case

In this section we introduce the more general definition of λ-correlation and prove the more general extension of the CH-bounds.

The proof of Theorem 1 is based on the observation that if we can upper bound each term $E[X^k]$ of the Taylor expansion of $E[e^{tX}]$ by $\lambda\ E[\hat{X}^k]$ where $\hat{X}$ is the sum of independent random variables, and if $E[e^{t\hat{X}}] \leq B$, then $E[e^{tX}] \leq \lambda\ B$. This motivates the following definition.

Definition 2 *Let $X_1, X_2, \ldots, X_n$ be bounded random variables such that $X_i \in [a_i, b_i]$ and let $X = \sum_{i \in [n]} X_i$. The X_i's are λ-correlated if there exists a collection of independent twin random variables $\hat{X}_i \in [a_i, b_i]$ such that,*

- *$E[X] \leq E[\hat{X}]$, where $\hat{X} = \sum_{i \in [n]} \hat{X}_i$, and*

- *for all nonempty $I \subseteq [n]$ and positive integers $s_1, s_2, \ldots, s_{|I|}$,*

$$E[\prod_{i\in I} X_i^{s_i}] \leq \lambda \prod_{i\in I} E[\hat{X}_i^{s_i}].$$

If we apply this definition to the case of 0-1 random variables we see that to have λ–correlated 0-1 random variables, it suffices to find twin variables $\hat{X}_i$ such that $E[X] \leq E[\hat{X}]$ and, for all nonempty $I \subseteq [n]$,

$$Pr\left(\bigwedge_{i\in I} X_i = 1\right) \leq \lambda \prod_{i\in I} Pr(\hat{X}_i = 1)$$

which reduces to definition 1 when $Pr(X_i = 1) = Pr(\hat{X}_i = 1)$. Notice that for this to hold, it is not necessary that $Pr(X_i = 1) = Pr(\hat{X}_i = 1)$; in typical applications we do not know the exact value of $Pr(X_i = 1)$ but only an approximation (usually an upper bound; we will see examples in later sections), but this is good enough to apply the CH-bounds extension.

Theorem 2 *Let X be the sum of λ-correlated random variables $X_1, X_2, \ldots, X_n$, where $X_i \in [a_i, b_i]$ and let $\hat{X}$ be the sum of the n twin variables $\hat{X}_i$. Then,*

$$Pr(X > (1+\epsilon)\, E[\hat{X}]) \leq \lambda\, G(\mu, \epsilon, \vec{a}, \vec{b}).$$

Proof. Let $\mu = E[\hat{X}]$. As in the proof of Theorem 1, we start with the inequality

$$\begin{aligned} Pr(X > (1+\epsilon)\mu) &= Pr(e^{tX} > e^{t(1+\epsilon)\mu}) \\ &\leq \frac{E[e^{tX}]}{e^{t(1+\epsilon)\mu}}. \end{aligned}$$

By the hypotheses of boundedness and of λ–correlation it follows that

$$\begin{aligned} E[e^{tX}] &= E\left[\sum_{k=0}^{\infty} \frac{t^k X^k}{k!}\right] \\ &= \sum_{k=0}^{\infty} \frac{t^k E[X^k]}{k!} \\ &\leq \lambda \sum_{k=0}^{\infty} \frac{t^k E[\hat{X}^k]}{k!} \\ &= \lambda\, E[e^{t\hat{X}}] \end{aligned}$$

By the already discussed result of Hoeffding [5], when $\hat{X}$ is the sum of n independent bounded random variables $\hat{X}_i \in [a_i, b_i]$

$$\min_{t>0} \frac{E[e^{t\hat{X}}]}{e^{t(1+\epsilon)\mu}} \leq G(n, \epsilon, \vec{a}, \vec{b}).$$

□

In this paper we will use the special case of Theorem 2 where $X_i \in [0,1]$, $i \in [n]$. In this case, $F(\mu,\epsilon)$ is also an upper bound for the upper tail of X.

Corollary 1 *Let X be the sum of n λ-correlated random variables $X_i \in [0,1]$. Then,*

$$Pr(X > (1+\epsilon)E[\hat{X}]) \leq \lambda\ F(\mu,\epsilon).$$

PROOF. Let $E[\hat{X}] = \mu$. When $\hat{X}$ is the sum of n independent random variables $\hat{X}_i \in [0,1]$ Hoeffding shows that, for $t \in (0, 1-\mu/n)$, (cf. Theorem 1 of [5])

$$Pr\left(\frac{\hat{X}-\mu}{n} \geq t\right) \leq \left(\frac{\mu}{\mu+nt}\right)^{\mu+nt}\left(1+\frac{n\,t}{n-\mu-nt}\right)^{n-\mu-nt}.$$

By setting $\epsilon = nt/n$ and by applying the standard approximation $1+x \leq e^x$ for $x = nt/(n-\mu-nt)$ we get

$$Pr(X \geq (1+\epsilon)\mu) \leq \left(\frac{e^{\epsilon}}{(1+\epsilon)^{(1+\epsilon)}}\right)^{\mu} = F(\mu,\epsilon).$$

□

4 The Edge Coloring Algorithm

We now present our randomized distributed edge coloring algorithm. The algorithm uses an idea of Karloff & Shmoys to reduce the problem of edge coloring general graphs to that of edge coloring a special class of bipartite graphs [7].

The Karloff & Shmoys algorithm was proposed for the PRAM model and is as follows. The input to the algorithm is a graph $G = (V,E)$ of maximum degree $\Delta(G) = \Delta$.

1. Compute a random partition of V into black and white vertices (all vertices flip a fair coin independently and in parallel). Let $G[B]$ be the subgraph induced by the black vertices, $G[W]$ be the subgraph induced by the white vertices, and $G[B,W]$ be the bipartite subgraph formed by the edges having endpoints of different colors.

2. Optimally edge color $G[B,W]$, *i.e.*, with $\Delta(G[B,W])$ colors, using the PRAM algorithm of Lev, Pippenger & Valiant [8].

3. Recurse on $G[B]$ and $G[W]$ using the same set of fresh new colors on both graphs.

The key fact is that, as long as Δ is large enough, the maximum degree of the three subgraphs is at most $\Delta/2 + \Delta^{1/2+\epsilon} = \Delta/2 + o(\Delta)$ *w.h.p.*, for any fixed $\epsilon > 0$. (This can be proved by applying the standard CH-bounds.) Note that the *same* set of fresh new colors can be used on both $G[B]$ and $G[W]$ because $G[B,W]$ completely separates the two graphs. The recurrence for the number of colors used is

$$\begin{aligned} C(\Delta) &\leq \left(\frac{\Delta}{2}+\Delta^{1/2+\epsilon}\right)+C\left(\frac{\Delta}{2}+\Delta^{1/2+\epsilon}\right) \\ &\leq \Delta+o(\Delta) \end{aligned}$$

and holds *w.h.p.* The "high probability" statement holds as long as $\Delta = \Omega(\log^{1+\delta} n)$, for any fixed $\delta > 0$. In this case, the failure probability given by the CH-bounds is the inverse of a superpolynomial function. For the case when the degree goes below the $\Omega(\log^{1+\delta})$ threshold, Karloff & Shmoys give a PRAM algorithm to $(\Delta+1)$–edge color graphs of this low maximum degree [7].

The non–distributed parts of the Karloff & Shmoys scheme are the subroutine which colors the bipartite graph optimally (it can be shown that this is impossible in the distributed model of computation [12]), and the handling of the case $\Delta \leq \log^{1+\delta} n$. The latter can be handled by Luby's randomized distributed $(\Delta+1)$–*vertex* coloring algorithm [9]. Luby's algorithm is simulated on the line graph of the given graph G thus giving a $2\Delta(G)-1$ edge coloring[2]. Hence, if we could get a good distributed algorithm for bipartite edge coloring then we could use the Karloff & Shmoys scheme to get a good coloring of any graph. We now show how this can be achieved.

4.1 Distributed Edge Coloring of Bipartite Graphs

We now present a simple Monte Carlo algorithm for edge coloring the special type of bipartite graphs generated by the Karloff & Shmoys scheme.

Given a bipartite graph $G=(A,B,E)$ we assume that each vertex knows whether it belongs to A or B. This is an important assumption because such information cannot be computed fast distributively [12], but it is verified by the bipartite graphs generated by the Karloff & Shmoys scheme. From now on, we will refer to vertices in A as the *top* vertices and to the vertices in B as the *bottom* vertices.

The algorithm takes $O(\log n)$ time and colors a bipartite graph G of maximum degree Δ with at most $1.6\Delta+16\log^{2+\delta} n$ colors *w.h.p.*, for any $\delta > 0$, which is approximately $1.6\Delta+\log^2 n$ (the failure probability will depend on δ). In the algorithm we use a variable Δ_C that is initialized to $\Delta(G)$. During any iteration of the algorithm, Δ_C is meant to be an upper bound on the degree of the current graph; we will prove later that this holds *w.h.p.* Recall that $\delta(u)$ denotes the set of edges incident on vertex u, and let THRESHOLD $= \log^{2+\delta} n$.

The algorithm is given two parameters $\epsilon, \delta > 0$, and is as follows:

1. PART I: Repeat until $\Delta_C <$ THRESHOLD:
   Let G_C be the current graph. Pick a set χ of Δ_C fresh new colors.

[2]When $\Delta(G) = O(polylog(n))$ we can compute a $2\Delta(G)-1$ colorings *deterministically* in $O(polylog(n))$ time with an algorithm based on the idea of removing maximal matchings. We prefer to use Luby's algorithm here for conciseness.

- *(Random proposal of bottom vertices)* In parallel and independently of the other vertices in B, each vertex $v \in B$ assigns a temporary color to each edge in $\delta(v)$ with uniform probability *without replacement, i.e.* edge e_1 is assigned color $\alpha \in \chi$ with probability $1/\Delta_C$, e_2 is assigned $\beta \in \chi - \{\alpha\}$ with probability $1/(\Delta_C - 1)$ and so on.
- *(Lottery of top vertices)* (Comment: The coloring so far is consistent around any vertex $v \in B$ but can be inconsistent around a vertex $u \in A$.) For each $u \in A$, let $C_u(\alpha)$ be the set of edges in $\delta(u)$ with temporary color α. Each vertex $u \in A$ selects a *winner* uniformly at random in $C_u(\alpha)$, for each nonempty $C_u(\alpha)$. All other edges, the *losers*, are decolored and assigned $\perp$.
- Set $\Delta_C := \Delta_C(1+\epsilon)/e$. $G_\perp$, the subgraph of G_C induced by the losers (*i.e.*, by the $\perp$-edges), becomes the new current graph.

2. PART II: Let G_r be the remaining graph. Edge color G_r with $2\Delta(G_r) - 1 < 2$ THRESHOLD colors by executing Luby's vertex coloring algorithm on the line graph of G_r [9].

The key claim is that in every iteration of part (I) above, the maximum degree of the graph shrinks by a factor of at least $(1+\epsilon)/e$ *w.h.p.*, as long as $\Delta \geq$ THRESHOLD. That is

$$\Delta(G_\perp) \leq (1+\epsilon)\,\frac{\Delta(G_C)}{e}$$

with high probability, for any fixed $\epsilon > 0$. The condition $\Delta_C \geq$ THRESHOLD ensures that the failure probability given by the extension of the CH-bounds is the inverse of a superpolynomial function. The reason for setting THRESHOLD $= \log^{2+\delta} n$ will be apparent from the probabilistic analysis.

Once the key claim is established, we can bound both the total number of colors used and the running time of the algorithm. To bound the number of colors used observe that if the degree of the graph shrinks at every iteration by at least a $(1+\epsilon)/e$ factor *w.h.p.* then the maximum degree of G_r is at most $\log^{2+\delta} n$ *w.h.p.* Hence, *w.h.p.*, the number of colors used is at most

$$C(\Delta) \leq \Delta + \frac{\Delta}{e}(1+\epsilon) + \ldots + \frac{\Delta}{e^k}(1+\epsilon)^k + 2\log^{2+\delta} n,$$

where k is the smallest integer such that $\Delta(1+\epsilon)^k/e^k \leq \log^{2+\delta} n$. The total number of colors used is at most (here, ϵ' depends on ϵ and can be made arbitrarily small)

$$\begin{aligned} C(\Delta) &\leq \left(\frac{e}{e-1} + \epsilon'\right)\Delta + \left(2 - \frac{e}{e-1} - \epsilon'\right)\log^{2+\delta} n \\ &< 1.585\,\Delta + 0.4\,\log^{2+\delta} n \\ &< 1.59\,\Delta \end{aligned}$$

when $\Delta > 8\,\log^{2+\delta} n$. The running time of the algorithm is $O(\log n)$ because part (I) takes $O(\log \Delta)$ time (by the key claim), and part (II), *i.e.*, Luby's algorithm, takes $O(\log n)$ time.

If $\Delta > 8\,\log^{2+\delta} n$ and we use our distributed subroutine for bipartite graphs in the Karloff & Shmoys algorithm the total number of colors is given by the recurrence

$$\begin{aligned} T(\Delta) &\leq C\left(\frac{\Delta}{2}+\Delta^{1/2+\epsilon}\right)+T\left(\frac{\Delta}{2}+\Delta^{1/2+\epsilon}\right) \\ &< 1.59\,\Delta+o(\Delta) \\ &< 1.6\,\Delta\,. \end{aligned}$$

If $\Delta \leq 8\ \log^{2+\delta} n$, we apply Luby's algorithm directly to get an edge coloring with $2\,\Delta-1 \leq 16\ \log^{2+\delta} n$. Hence, the total number of colors to color *any* graph is at most $1.6\,\Delta+16\ \log^{2+\delta} n$, for any $\delta > 0$, which is approximately $1.6\,\Delta + \log^2 n$ colors.

The rest of the paper is devoted to establishing the key claim.

5 Probabilistic Analysis

This section is devoted to proving the key claim of the preceding section, namely that given a graph G and Δ such that $\Delta \geq \Delta(G)$ and $\Delta \geq$ THRESHOLD then, after one iteration of Part (I) of the bipartite algorithm, the maximum degree of the new graph, $\Delta(G_\perp)$, is at most $(1+\epsilon)\Delta/e$ *w.h.p.*, for any fixed $\epsilon > 0$.

It turns out that the analysis is considerably easier for the top vertices than for the bottom vertices. We begin with the easy part.

5.1 Analysis of Top Vertices

Let u be a generic top vertex with neighbors v_i, $i \in [d_u]$, and incident edges $i = (u, v_i)$. Recall that a *loser* is an edge which, after having got a tentative color in the random proposal, lost the lottery and got decolored. So, the new degree of u is given by the number of losers incident with it.

From the point of view of a top vertex, the random proposal and the lottery are equivalent to the following random experiment. For each edge i incident on u we introduce a ball i, and for each color k we introduce a bin k; the assignment of a tentative color to an edge by the algorithm is equivalent to throwing each ball into one of the Δ bins independently and uniformly at random, since the bottom vertices assign tentative colors with uniform probability and independently of the other bottom vertices. Recalling that we have at most Δ balls and exactly Δ bins:

$$\begin{aligned} \sharp\mathit{losers} &= \sharp\mathit{balls} - \sharp\mathit{winners} \\ &\leq \sharp\mathit{bins} - \sharp\mathit{nonempty\ bins} \\ &= \sharp\mathit{empty\ bins}. \end{aligned}$$

Let X be a random variable denoting the number of losers. To estimate X and its tail distribution we will study the random variable $B = \sharp\mathit{empty\ bins}$. For this purpose, we introduce Δ many indicator random variables

$$B_i = \begin{cases} 1 & \text{bin } i \text{ is empty} \\ 0 & \text{otherwise} \end{cases}$$

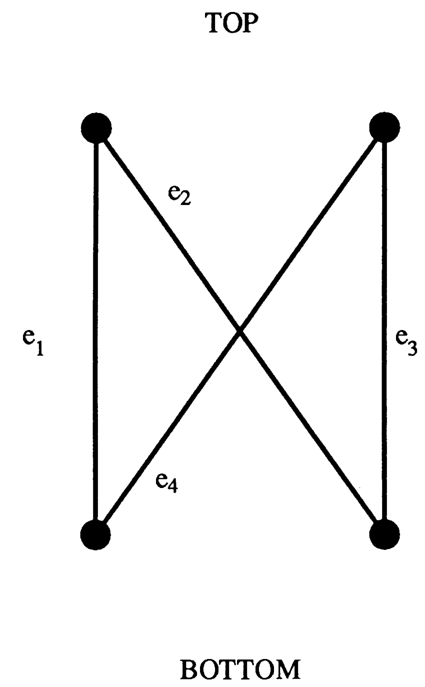

Figure 1: $\chi(e_1) = \alpha$, $\chi(e_2) = \beta$, e_1 lost the lottery

and hence, $B = \sum_{i \in [\Delta]} B_i$. Notice that $X \leq B$ always. The variable B was studied in Section 3 where it was shown that $E[B] \leq \Delta/e$ and that the B_i's are 1-correlated, which implies that $Pr(B > (1+\epsilon)\Delta/e) \leq F(\Delta/e, \epsilon)$. Since $E[X] \leq E[B]$ and $Pr(X > (1+\epsilon)\Delta/e) \leq Pr(B > (1+\epsilon)\Delta/e)$ we get

Theorem 3 *Let u be a top vertex and X be the random variable denoting the number of losers incident on it. Then, $E[X] \leq \Delta/e$ and*

$$Pr(X > (1+\epsilon)\Delta/e) \leq F(\Delta/e, \epsilon).$$

5.2 Analysis of Bottom Vertices

In this section we analyze what happens to the new degree of a generic bottom vertex v_B. This case is considerably harder to handle than the previous one, because of the way in which the random variables describing the process are correlated. For a top vertex, the dependency among the variables was playing for us; given that some edges incident on a top vertex are losers, the probability of having another loser decreases. For a bottom vertex the situation is reversed: having some edges losing the lottery might even make the probability of having another loser increase. The problem can be seen in Figure 1. Suppose we are given that e_1 got tentative color α and lost the lottery, and that e_2 got tentative color β; we will argue intuitively that given this, the probability of e_2 losing the lottery has increased. Since e_1 lost, the probability of e_3 getting tentative color α increases, which implies that the probability of e_4 getting tentative color β also increases, and this increases the probability of e_2 losing the lottery.

For the sake of the analysis we modify the algorithm as follows: instead of performing all random proposals in parallel, suppose that the bottom vertices perform their random proposals sequentially, one after the other. This does not modify the probability distributions because the choices are still done independently. We want to focus our attention on the last vertex v_B performing the random proposal. We will use the fact that when v_B performs its random proposal, all edges not incident on v_B already have a tentative color. By symmetry, any upper bound on the probabilities we can find for v_B will hold for all bottom vertices.

Let $i \in [d_v]$ denote an edge incident with the bottom vertex v_B, with the other endpoint of i being u_i. We introduce the indicator random variables

$$X_i = \begin{cases} 1 & i \text{ loses the lottery} \\ 0 & \text{otherwise} \end{cases}$$

and want to study the expectation and tail probability distribution of $X = \sum_{i \in [d_v]} X_i$. Computing the expectation is easy.

Lemma 1 $E[X] \leq \Delta/e$.

Proof. Let v_B be the bottom vertex. It is sufficient to show that $Pr(X_i = 1) \leq 1/e$, for all $i \in [d_v]$. From the analysis of the top vertices, we know that the expected number of losers incident with u_i is at most Δ/e and hence that $\sum_{j \in \delta(u_i)} Pr(j \text{ loses}) \leq \Delta/e$. By symmetry, $Pr(j) \leq 1/e$ for all $j \in \delta(u_i)$, and hence $Pr(X_i = 1) \leq 1/e$. □

We now study the tail probability distribution of X. Our goal is to show that $X \leq (1+\epsilon)\Delta/e$ *w.h.p.*, for any fixed $\epsilon > 0$. Establishing this claim will take several lemmas.

For technical reasons that will be clear later, we subdivide the edges incident on v_B into $\sqrt{\Delta}$ groups, each of at most $\sqrt{\Delta}$ edges. Let G be any such group and define $X_G = \sum_{i \in G} X_i$; we would like to show that X_G is the sum of λ-correlated random variables, so that

$$\forall \epsilon > 0, \ \ Pr(X_G > (1+\epsilon)\sqrt{\Delta}/e) \leq \lambda \, F(\sqrt{\Delta}/e, \epsilon).$$

Rather than proving this claim, we will prove a weaker one namely that X_G is the sum of λ-correlated random variables *with high probability!* More precisely, we will define an event $\mathcal{A}$ such that $\mathcal{A}$ happens *w.h.p.*, and such that, given that $\mathcal{A}$ occurs, then X_G is the sum of λ-correlated random variables. That is,

$$Pr(X_G > (1+\epsilon)\sqrt{\Delta}/e \mid \mathcal{A}) \leq \lambda \, F(\sqrt{\Delta}/e, \epsilon).$$

Showing this weaker claim is sufficient because

$$\begin{aligned} Pr(X > (1+\epsilon)\Delta/e) &\leq Pr(\exists X_G > (1+\epsilon)\sqrt{\Delta}/e) \\ &= Pr(\exists G \ X_G > (1+\epsilon)\sqrt{\Delta}/e \mid \mathcal{A}) \ Pr(\mathcal{A}) \ + \\ &\quad\ Pr(\exists G \ X_G > (1+\epsilon)\sqrt{\Delta}/e \mid \mathcal{A}^c) \ Pr(\mathcal{A}^c) \\ &\leq Pr(\exists G \ X_G > (1+\epsilon)\sqrt{\Delta}/e \mid \mathcal{A}) \ + \ Pr(\mathcal{A}^c) \\ &\leq \sqrt{\Delta} \ Pr(X_G > (1+\epsilon)\sqrt{\Delta}/e \mid \mathcal{A}) \ + \ Pr(\mathcal{A}^c) \\ &\leq \sqrt{\Delta} \ \lambda \ F(\sqrt{\Delta}/e, \epsilon) \ + \ Pr(\mathcal{A}^c). \end{aligned}$$

Another interesting point in the analysis is that λ will not be a constant term but an exponentially growing function. However we will be able to show the rate of growth of λ is much slower than that of $1/F(\sqrt{\Delta}/e, \epsilon)$. In other words, we will show that $\lambda\ F(\sqrt{\Delta}/e, \epsilon)$ goes to zero superpolynomially fast.

We now turn to the task of showing that X_G is the sum of λ-correlated random variables *w.h.p.* Recall from the definition that the X_i's, $i \in G$, are λ-correlated if, for all nonempty $I \subseteq G$,

$$Pr\left(\bigwedge_{i\in I} X_i = 1\right) \leq \lambda \prod_{i\in I} Pr(X_i = 1).$$

Consider then a generic subset $I \subseteq G$ of edges incident on the bottom vertex v_B, and let us see how to compute $Pr\left(\bigwedge_{i\in I} X_i = 1\right)$. Recall that we are analyzing the situation where v_B is the last vertex to perform its random proposal. This means that prior to the assignment of a tentative color to edge $i = (v, u_i)$, all other edges incident on u_i already have their tentative color. Using the balls-and-bins language, we can say that prior to throwing ball i at random into one of the bins, all balls coming from the other neighbors of u_i have been thrown. We will think of i as a red ball and of the other edges at u_i as white balls. Once the red ball is thrown in, say, bin $k \in [\Delta]$, a winner is selected uniformly at random among all (*i.e.*, red *and* white) balls in bin k. All other balls, the losers, are discarded. Notice that the probability of discarding the red ball is itself a random variable which depends on the particular placement of the white balls prior to throwing the red ball.

Given any placement of white balls at u_i, we construct a vector of probabilities C_i as follows. Let a_{ik} denote the number of white balls in bin $k \in [\Delta]$ of vertex u_i, and let $p_{ik} = a_{ik}/(1 + a_{ik})$ denote the probability that the red ball loses the lottery given that it was thrown in bin k (equivalently, p_{ik} is the probability that edge i loses, given that it got tentative color k). For each neighbor u_i of our bottom vertex v_B, we construct the corresponding vector $C_i = (p_{i1}, p_{i2}, \ldots, p_{i\Delta})$. Given a set I of neighbors of v_B we construct the matrix M_I whose i-th column is the vector C_i. The next lemma explains why this matrix is relevant to us. From now on, let $p(m,k) \doteq m(m-1)\cdots(m-k+1)$.

Lemma 2 *Let $I \subseteq G$ and $k = |I|$. Then,*

$$Pr\left(\bigwedge_{i\in I} X_i = 1\right) = \frac{perm(M_I)}{p(\Delta, k)}.$$

PROOF. The random proposal of v_B is equivalent to choosing an one-to-one function $\pi : I \to [\Delta]$ uniformly at random among the set $\mathcal{P}$ of all such functions. Recall that the entry M_{ik} of M_I is the probability p_{ik} that edge i loses given that it is given color k. Hence,

$$\begin{aligned}
Pr(\wedge_{i\in I} X_i = 1) &= \sum_{\pi\in\mathcal{P}} Pr(\wedge_{i\in I} X_i = 1 \mid \pi \textit{ is selected})\ Pr(\pi \textit{ is selected}) \\
&= \sum_{\pi\in\mathcal{P}} \left(M_{1,\pi(1)}\ M_{2,\pi(2)}\ \cdots\ M_{k,\pi(k)}\right)\left(\frac{1}{\Delta}\cdot\frac{1}{\Delta-1}\cdot\cdots\cdot\frac{1}{\Delta-k+1}\right) \\
&= \frac{perm(M_I)}{p(\Delta, k)}.
\end{aligned}$$

□

We now want to find a good upper bound of $perm(M_I)$. The following lemma gives a simple upper bound that is sufficient for our purposes. Given a matrix M let S_i denote the sum of the elements of C_i, the i-th column of M.

Lemma 3 *Let $M = [M_{ik}]$ be a matrix with c columns and r rows, and non-negative entries M_{ik}. Then, $perm(M) \leq \prod_{i \in [c]} S_i$.*

PROOF. Let $\mathcal{P} = \{\pi \mid \pi : [c] \rightarrow [r],\ \pi \text{ is one-to-one}\}$. Then,

$$\begin{aligned} perm(M) &= \sum_{\pi \in \mathcal{P}} M_{1,\pi(1)}\, M_{2,\pi(2)}\, \cdots\, M_{c,\pi(c)} \\ &\leq (M_{1,1} + \ldots + M_{1,r})(M_{2,1} + \ldots + M_{2,r}) \ldots (M_{c,1} + \ldots + M_{c,r}) \\ &= \prod_{i \in [c]} S_i. \end{aligned}$$

□

The next lemma relates the value of S_i to that of $1/e \geq Pr(i \text{ loses})$, $i \in \delta(v)$. It is an application of the most general definition of λ–correlation. Before the proof of the lemma, we establish

Proposition 1 *If $0 \leq p \leq 1$, $q = 1 - p$ and m is a positive integer, then*

$$\sum_{r=1}^{m} \binom{m}{r} p^r q^{m-r} \frac{r}{r+1} = 1 - \frac{(1 - q^{m+1})}{p(m+1)}.$$

PROOF. Let

$$\begin{aligned} f(p) &= \sum_{r=1}^{m} \binom{m}{r} p^r q^{m-r} \frac{k}{k+1} \\ &= 1 - q^m - \sum_{r=1}^{m} \binom{m}{r} p^r q^{m-r} \frac{1}{k+1}. \end{aligned}$$

Integrating both sides of the binomial expansion

$$(x+q)^m = \sum_{r=0}^{m} \binom{m}{r} x^r q^{m-r}$$

between 0 and p, we get

$$\frac{1 - q^{m+1}}{m+1} = p\,(1 - f(p)),$$

from which the proposition follows. □

We now return to our scenario where v_B is the last bottom vertex to pick tentative colors for its edges, with its edges having been split into groups. Recall that we are focusing on a particular group G, and that we want a good upper bound on $Pr\left(\bigwedge_{i \in I} X_i = 1\right)$, for any nonempty

$I \subseteq G$. Lemma 2 gives the upper-bound $Pr\left(\bigwedge_{i\in I} X_i = 1\right) \leq perm(M_I)/p(\Delta, |I|)$, and Lemma 3 gives the upper-bound $perm(M_I) \leq \Pi S_i$. So, a good upper-bound of S_i will hopefully translate into a good upper-bound for $Pr\left(\bigwedge_{i\in I} X_i = 1\right)$. The next lemma says that $S_i \leq \Delta/e$ *w.h.p.* Recall that we are analyzing the situation where v_B is to give tentative colors for its edges and that all other edges have alraedy been assigned tentative colors. If we focus on a particular edge $i = (v_B, u_i)$ the situation is equivalent to throwing a red ball i uniformly at random into Δ bins where each bin has some number of white balls in it (possibly zero). Notice that the number of white balls in each bin is itself a random variable.

Lemma 4 *Let $i = (v, u_i)$ be any edge in $I \subseteq G$, and S_i be the sum of the elements of C_i, the i-th column of M_I. Then, $E[S_i] \leq \Delta/e = \mu$ and*

$$\forall \epsilon_1 > 0, \;\; Pr(S_i > (1+\epsilon_1)\mu) \leq F(\mu, \epsilon_1).$$

Proof. Let Z_k be the random variable denoting the number of white balls in bin k and $Y_k = Z_k/(Z_k+1)$ be the random variable denoting the probability that the red ball loses the lottery given that it ends in bin k. Then, $S_i = Y \doteq \sum_k Y_k$. Note that the Y_k's are bounded random variables with values in $[0,1]$. We will show that $E[Y] \leq \Delta/e$ and that the Y_k's are 1–correlated (under the most general definition of λ–correlation), which will give our claim.

Firstly, we may assume that we have $d = \Delta - 1$ white balls (*i.e.*, the degree of u_i is Δ): $Pr(Y \geq (1+\epsilon_1)\Delta/e)$ is maximized at $d = \Delta - 1$, as d varies from 1 to $\Delta - 1$. (To see this, assume $d = \Delta - 1 - k < \Delta - 1$. Add k *yellow* balls to the white balls, and run two experiments. In one experiment throw the white and red balls and compute the probability that the red ball loses the lottery. In the other experiment, throw white, yellow and red balls and again compute the probability that the red ball loses. In both experiments, let us look at the bin where the red ball fell. The probability that the red ball loses is $b/(b+1)$ for the first experiment, and $(b+y)/(b+y+1)$ for the second, where b and y are, respectively, the number of white and yellow balls in the bin. Since $y \geq 0$, $b/(b+1) \leq (b+y)/(b+y+1)$. If $Y_i(d)$ indicates the variable Y_i when u_i has degree d, then $Y_i(d) \leq Y_i(\Delta)$ for all $i \in [\Delta]$ and $d \in [\Delta]$.)

First, we will show that, for all i, $E[Y_i] \leq 1/e$ and then we will show that, for any set of k indices $I \subseteq [\Delta]$ and strictly positive integers s_i,

$$E[\prod_{i\in I} Y_i^{s_i}] \leq 1/e^k. \tag{3}$$

Given this we can apply Corollary 1 by introducing n independent twin 0-1 random variables $\hat{Y}_i$ such that $E[\hat{Y}_i] = Pr(\hat{Y}_i = 1) = 1/e$. Since the $\hat{Y}_i$'s are binary, equation 3 is the same as

$$E[\prod_{i\in I} Y_i^{s_i}] \leq \prod_{i\in I} E[\hat{Y}_i] = \prod_{i\in I} E[\hat{Y}_i^{s_i}],$$

which is to say that the Y_i's are 1–correlated. Noting that $0 \leq Y_i \leq 1$, it suffices to show that

$$E[\prod_{i\in I} Y_i] \leq 1/e^k. \tag{4}$$

Without loss of generality we can assume $I = [k]$. We will prove inequality 4 by induction on $k \geq 1$; when $k = 1$,

$$\begin{aligned} E[Y_1] &= \sum_{r=0}^{\Delta-1} \binom{\Delta-1}{r} \left(\frac{1}{\Delta}\right)^r \left(1-\frac{1}{\Delta}\right)^{\Delta-1-r} \frac{r}{r+1} \\ &= (1-1/\Delta)^{\Delta} \leq 1/e, \end{aligned}$$

where the second equality follows from Proposition 1. Notice that for all $j \in [\Delta]$, $E[Y_j] = E[Y_1] \leq 1/e$. When $k > 1$, the law of conditional probabilities gives

$$E[\prod_{i\in[k]} Y_i] = E[Y_1Y_2\cdots Y_{k-1}\; E[Y_k \mid Y_1Y_2\cdots Y_{k-1}]] \tag{5}$$

Suppose we show that for all *non-zero* $c_i \in [0,1]$ with $i \in [k-1]$,

$$E[Y_k \mid \bigwedge_{i=1}^{k-1} Y_i = c_i] \leq \frac{1}{e}; \tag{6}$$

then, since the product $Y_1Y_2\cdots Y_{k-1}$ in equation (5) is zero when any c_i is zero, we see by induction on k that

$$\begin{aligned} E[\prod_{i=1}^{k} Y_i] &= E[Y_1Y_2\cdots Y_{k-1}\; E[Y_k \mid Y_1Y_2\cdots Y_{k-1}]] \\ &\leq \frac{1}{e} E[\prod_{i=1}^{k-1} Y_i] \\ &\leq \frac{1}{e^k}. \end{aligned}$$

Hence, the claim follows if we can show that inequality (6) holds.

If a_i denotes the number of white balls that fell into bin i, then $c_i = a_i/(a_i+1)$. Let $a = \sum_{i=1}^{k-1} a_i \geq k-1$, $p = 1/(\Delta - k + 1)$, and $q = 1-p$. Then

$$\begin{aligned} E[Y_k| \bigwedge_{i=1}^{k-1} Y_i = c_i] &= E[Y_k| \bigwedge_{i=1}^{k-1} Z_i = a_i] \\ &= \sum_{r=1}^{\Delta-1-a} t(r,a), \end{aligned}$$

where

$$t(r,a) \doteq \binom{\Delta-1-a}{r} p^r q^{\Delta-1-a-r} \frac{r}{r+1}.$$

It is easy to check that $t(r,a) \geq t(r,a+1)$. As a consequence, the maximum value of $E[Y_k| \bigwedge_{i=1}^{k-1} Y_i = c_i]$ is attained at $a = k-1$, in which case we have

$$\begin{aligned}
\sum_{r=1}^{\Delta-1-a} t(r,a) &= \sum_{r=1}^{\Delta-k} t(r,k-1) \\
&= \sum_{r=1}^{\Delta-k} \binom{\Delta-k}{r} p^r q^{\Delta-k-r} \frac{r}{r+1} \\
&= q^{\Delta-k+1} \leq 1/e,
\end{aligned}$$

by Proposition 1. □

Putting all the pieces together (recall that $k = |I|$):

$$\begin{aligned}
Pr(\wedge_{i \in I} X_i = 1) &= \frac{\mathrm{perm}(M_I)}{p(\Delta,k)} \\
&\leq \frac{\prod_{i \in I} S_i}{p(\Delta,k)} \\
&\leq (1+\epsilon_1)^k \frac{\Delta^k}{e^k} \frac{1}{p(\Delta,k)} \\
&\leq (1+\epsilon_1)^k \frac{\Delta^k}{p(\Delta,k)} \frac{1}{e^k},
\end{aligned}$$

for any $\epsilon_1 > 0$. The first line follows from Lemma 2, the second from Lemma 3, and the third from Lemma 4 and holds *w.h.p.* The last line is just a re-writing of the third one. Given $\epsilon_1 > 0$, let $\mathcal{A}_{\epsilon_1}$ be the event that, for all $i \in G$, $S_i \leq (1+\epsilon_1)\Delta/e$. Lemma 4 says that $\mathcal{A}_{\epsilon_1}$ happens *w.h.p.*, for any $\epsilon_1 > 0$. The last line of the above chain of inequalities means that, given that $\mathcal{A}_{\epsilon_1}$ occurred, the X_i's are λ-correlated with $\lambda = (1+\epsilon_1)^k \Delta^k / p(\Delta,k)$. (Alternatively, we can say that the X_i's are λ–correlated *w.h.p.*) We now want an estimate of λ. The next lemma explains why we subdivided the edges incident on v_B into groups.

Lemma 5 *For positive integers t and k, $t^k/p(t,k) \leq e^{k^2/t}$, if $k \leq t/2$.*

Proof. We first note that $\ln(1-x) \geq -2x$, for $0 \leq x \leq 1/2$. This is true, since if we define $f(x) \doteq \ln(1-x) + 2x$, then $f(0) = 0$ and $f'(x) = \frac{1-2x}{1-x}$, which is non-negative for $0 \leq x \leq 1/2$. Now,

$$\begin{aligned}
\frac{p(t,k)}{t^k} &= \prod_{i=1}^{k-1} \frac{(t-i)}{t} \\
&= \exp\left(\sum_{i=1}^{k-1} \ln(1-\frac{i}{t})\right) \\
&\geq \exp\left(-\sum_{i=1}^{k-1} \frac{2i}{t}\right) \qquad \text{(since } k \leq t/2\text{)} \\
&= \exp\left(-\frac{(k-1)k}{t}\right) \\
&\geq e^{-\frac{k^2}{t}}.
\end{aligned}$$

$\square$

For $t = \Delta$ and $k = |I| \leq \sqrt{\Delta}$ this lemma implies that $\Delta^k/p(\Delta, k) \leq e$ and hence, given $\mathcal{A}_{\epsilon_1}$, the X_i are λ-correlated with $\lambda = (1+\epsilon_1)^k e \leq e^{1+\epsilon_1\sqrt{\Delta}}$, via the standard approximation $1 + x \leq e^x$. Though λ goes to infinity superpolynomially fast as Δ goes to infinity, we can still upper bound $Pr(X_G > (1+\epsilon)\Delta/e)$ by a term that goes to zero superpolynomially fast, as follows. (Here, $\mu = \sqrt{\Delta}/e$)

$$\begin{aligned}
Pr(X_G > (1+\epsilon)\mu) &= Pr(X_G > (1+\epsilon)\mu \mid \mathcal{A}_{\epsilon_1})\, Pr(\mathcal{A}_{\epsilon_1}) + Pr(X_G > (1+\epsilon)\mu \mid \mathcal{A}^c_{\epsilon_1})\, Pr(\mathcal{A}^c_{\epsilon_1}) \\
&\leq Pr(X_G > (1+\epsilon)\mu \mid \mathcal{A}_{\epsilon_1}) + Pr(\mathcal{A}^c_{\epsilon_1}) \\
&\leq \lambda\, F(\mu, \epsilon) + Pr(\exists G\; X_G > (1+\epsilon_1)\mu) \\
&\leq e^{1+\epsilon_1\sqrt{\Delta}}\, F(\mu, \epsilon) + \sqrt{\Delta}\, F(\mu, \epsilon_1) \\
&\leq e^{1+\epsilon_1\sqrt{\Delta}}\, e^{-\epsilon^2\sqrt{\Delta}/3} + \sqrt{\Delta}\, F(\mu, \epsilon_1) \\
&\leq e^{1+(\epsilon_1-\epsilon^2/3)\sqrt{\Delta}} + \sqrt{\Delta}\, F(\mu, \epsilon_1).
\end{aligned}$$

Since ϵ_1 can be chosen arbitrarily small, we can make the above term go to zero exponentially fast by choosing ϵ_1 such that $\epsilon_1 - \epsilon^2/3 < 0$.

Remark. We can now see why the parameter THRESHOLD must be $\Omega(\log^{2+\delta} n)$: the above failure probability goes to zero superpolynomially fast as long as $\Delta = \Omega(\log^{2+\delta} n)$, for any fixed $\delta > 0$.

Hence, we have proven that $X_G < (1+\epsilon)\sqrt{\Delta}/e$ *w.h.p.*, for any fixed $\epsilon > 0$. In turn, this implies that the new degree of v_B after one iteration of the algorithm is at most $(1+\epsilon)\Delta/e$ *w.h.p.* To see this,

$$\begin{aligned}
Pr(X > (1+\epsilon)\Delta/e) &\leq Pr(\exists G\; X_G > (1+\epsilon)\sqrt{\Delta}/e) \\
&\leq \sqrt{\Delta}\, Pr(X_G > (1+\epsilon)\sqrt{\Delta}/e) \\
&\leq \sqrt{\Delta}\, \left(e^{1+(\epsilon_1-\epsilon^2/3)\sqrt{\Delta}} + \sqrt{\Delta} F(\mu, \epsilon_1)\right).
\end{aligned}$$

This inequality holds for any ϵ and ϵ_1. Hence the new degree of *all* the vertices is at most $(1+\epsilon)\Delta/e$ *w.h.p.* and we get

Theorem 4 *The new degree of the graph after one iteration of Part (I) of the algorithm is at most* $(1+\epsilon)\Delta/e$ *w.h.p., for any fixed* $\epsilon > 0$.

References

[1] B. Awerbuch, A. V. Goldberg, M. Luby, and S. A. Plotkin. Network decomposition and locality in distributed computation. In *Proceedings of the IEEE Symposium on Foundations of Computer Science*, pages 364–369, 1989.

[2] B. Berger and J. Rompel. Simulating ($\log^c n$)-wise independence in NC. *J. Assoc. Comput. Mach.*, 38(4):1026–1046, 1991.

[3] B. Bollobás. *Graph Theory.* Springer Verlag, New York, 1979.

[4] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23:493–509, 1952.

[5] W. Hoeffding. Probability inequalities for sums of bounded random variables. *American Statistical Association Journal*, pages 13–30, 1963.

[6] R. Jain, K. Somalwar, J. Werth, and J. C. Browne. Scheduling parallel i/o operations in multiple bus systems. *Journal of Parallel and Distributed Computing*, 16(4):352–362, 1992.

[7] H. J. Karloff and D. B. Shmoys. Efficient parallel algorithms for edge coloring problems. *Journal of Algorithms*, 8:39–52, 1987.

[8] G. F. Lev, N. Pippenger, and L. G. Valiant. A fast parallel algorithm for routing in permutation networks. *IEEE Transactions on Computers*, 30:93–100, 1981.

[9] M. Luby. Removing randomness in parallel computation without a processor penalty. In *Proceedings of the IEEE Symposium on Foundations of Computer Science*, pages 162–173, 1988.

[10] R. Motwani, J. Naor, and M. Naor. The probabilistic method yields deterministic parallel algorithms. In *Proceedings of the IEEE Symposium on Foundations of Computer Science*, pages 8–13, 1989.

[11] A. Panconesi and A. Srinivasan. Fast randomized algorithms for distributed edge coloring. In *Proceedings of the ACM Symposium on Principles of Distributed Computing*, pages 251–262, 1992.

[12] A. Panconesi and A. Srinivasan. Improved distributed algorithms for coloring and network decomposition problems. In *Proceedings of the ACM Symposium on Theory of Computing*, pages 581–592, 1992.

[13] P. Raghavan. Lecture notes on randomized algorithms. Technical Report RC 15340 (#68237), IBM T.J.Watson Research Center, January 1990. Also available as CS661 Lecture Notes, Technical report YALE/DCS/RR-757, Department of Computer Science, Yale University, January 1990.

[14] R. Raman. The power of Collision: Randomized parallel algorithms for chaining and integer sorting. In *Proceedings, 10th Annual FST & TCS Conference*, Lecture Notes in Computer Science # 472, pages 161–175. Springer-Verlag, Berlin, December 1990. Also available as University of Rochester CS Dept. TR 336, March 1990 (Revised January 1991).