Matchings in Superpositions of (n, n)-Bipartite Trees

Eric Schmutz*
Mathematics and Computer Science Department, Drexel University, Philadelphia, PA 19104

ABSTRACT

Pick two trees from among all "bipartite trees" with a fixed (n, n) two-coloring. We estimate the probability that the superposition of these two trees contains a perfect matching. As $n \to \infty$, this probability approaches 1. © 1994 John Wiley & Sons, Inc.

1. INTRODUCTION

Partition $[2n]$ into two sets of equal cardinality, say $L_n = \{l_1, l_2, \ldots, l_n\}$ and $D_n = \{d_1, d_2, \ldots, d_n\}$. Of the $(2n)^{2n-2}$ labelled trees on $[2n]$, exactly n^{2n-2} have the property that each edge joins a vertex in L_n to a vertex in D_n. (For a proof, see [6] and [4], or use Lemma 2 below). Independently choose r trees from among these n^{2n-2} "bipartite trees." Let $G = G(T_1, T_2, \ldots, T_r)$ be the graph on $[2n]$ that is obtained by taking the union of the edges of the r trees. We are interested in $P_r(n) \stackrel{\text{def}}{=}$ the probability that G contains a perfect matching. From the results in Walkup [8] and Palmer [5], one might expect that $P_2(n) \to 1$ as $n \to \infty$. The main result in this paper is:

Theorem 1. $1 - P_2(n) = O(\log n / n)$.

If K is a graph and S is a set of vertices in K, define $N_K(S)$ to be the set of vertices in K that are adjacent to one or more of the vertices in S. Define (S, S') to be a *bad* $(k, k-1)$ *pair* if $S \subseteq L_n$, $N_G(S) \subseteq S'$, $|S| = k$, and $|S'| = k - 1$. Similarly (S, S') is a *bad* $(k-1, k)$ *pair* if $S' \subseteq D_n$, $N_G(S') \subseteq S$, $|S'| = k$, and

* Research supported by NSF Grant DMS 9101753.

Random Structures and Algorithms, Vol. 5, No. 1 (1994)

$|S| = k - 1$. We say that (S, S') is a *bad k pair* if it is a bad $(k, k-1)$ pair or a bad $(k-1, k)$ pair, and we say that (S, S') is a bad pair if it is a bad k pair for some k.

Hall's theorem [3] asserts that G contains a perfect matching if and only if it contains no bad pair. Observe that (S, S') is a bad $(k, k-1)$ pair if and only if $(L_n - S, D_n - S')$ is a bad $(n-k, n-k+1)$ pair. Hence we need only check for bad k pairs with $k \leq n/2$. Let $\mathbf{B}_k(G)$ be the number of bad k-pairs, and let $\mathbf{B}(G) = \sum_{k=2}^{\lceil n/2 \rceil} \mathbf{B}_k(G)$. Then we have

Observation 1. *G has a perfect matching if and only if* $\mathbf{B}(G) = 0$.

Palmer [5] proved that, for $r \geq 3$, the average value of $\mathbf{B}$ approaches 0 as $n \to \infty$. It follows immediately that, for $r \geq 3$, $P_r(n) \to 1$. This first moment argument breaks down for $r = 2$; when $r = 2$, $E(\mathbf{B})$ is bounded away from zero.

Define (S, S') to be a *minimal bad* $(k, k-1)$ *pair* if

(1) (S, S') is a bad $(k, k-1)$ pair,
(2) there is no pair (A, A') such that $A \subset S$, $A' \subset S'$, $|A| < |S|$, and (A, A') is a bad $(|A|, |A| - 1)$ pair.

Minimal bad $(k-1, k)$ pairs and minimal bad k pairs are defined in a similar way. Let $\mathbf{M}_k(G)$ be the number of minimal bad k pairs, and let $M(G) = \sum_{k=2}^{\lceil n/2 \rceil} \mathbf{M}_k(G)$. The following is an immediate corollary to observation 1:

Observation 2. *G has a perfect matching if and only if* $\mathbf{M}(G) = 0$.

If we can prove that $E(\mathbf{M}) = O(\log n/n)$ as $n \to \infty$, then Theorem 1 will be proved. Observe that

$$\begin{aligned} E(\mathbf{M}) &= \sum_{S \subseteq L_n, S' \subseteq D_n} \text{Prob}((S, S') \text{ is a minimal bad } k \text{ pair for some } k \leq \lceil n/2 \rceil) \\ &= 2 \sum_{\substack{S \subseteq L_n, S' \subseteq D_n \\ |S'| \leq \lceil n/2 \rceil}} \text{Prob}((S, S') \text{ is a minimal bad } (|S'| - 1, |S'|) \text{ pair}) \\ &= 2 \sum_{k=2}^{\lceil n/2 \rceil} \sum_{\substack{S \subseteq L_n, |S| = k-1 \\ S' \subseteq D_n, |S'| = k}} \text{Prob}((S, S') \text{ is a minimal bad } (k-1, k) \text{ pair}). \end{aligned}$$

Note that the $\binom{n}{k-1}\binom{n}{k}$ terms of the inner sum are all equal. Let $L_{k-1} = \{l_1, \ldots, l_{k-1}\}$, let $D_k = \{d_1, \ldots, d_k\}$, and let $Q_k := \text{Prob}((L_{k-1}, D_k)$ is a minimal bad $(k-1, k)$ pair). Then we have

$$E(\mathbf{M}) = 2 \sum_{k=2}^{\lceil n/2 \rceil} \binom{n}{k-1} \binom{n}{k} Q_k . \tag{1}$$

We need to estimate Q_k.

If H is a subgraph of a graph K, and if v is a vertex of H, let $\rho_H(v)$ be the degree of v in H. In particular, let H_k be the subgraph of G that is induced by [illegible] D_k. Then for $v \in L_{k-1}$, $\rho_{H_k}(v)$ is the number of vertices in D_k that are [illegible] to v. Note that $\rho_{H_k}(v)$ may be strictly less than the degree of v in G. We [illegible]

Lemma 1. *If (L_{k-1}, D_k) is a minimal bad pair, and if $k \geq 3$, then for all $v \in L_{k-1}$, we have $\rho_{H_k}(v) \geq 2$.*

Proof. By contradiction. Suppose $v \in L_{k-1}$ and $\rho_{H_k}(v) < 2$. If $\rho_{H_k}(v) = 1$, then let w be the neighbor of v that is in D_k. Otherwise, let w be an arbitrary vertex in D_k. In either case, $(L_{k-1} - \{v\}, D_k - \{w\})$ is a bad $(k-2, k-1)$ pair, contradicting the minimality of (L_{k-1}, D_k). ∎

Now let $\mathbf{S}_k(G) := \sum_{j=1}^{k-1} \mathbf{I}_j(G)$, where $\mathbf{I}_j(G)$ is 1 if $\rho_{H_k}(l_j) = 0$, and zero otherwise. In other words, $\mathbf{S}_k(G)$ is the number of vertices in L_{k-1} that have no neighbors in D_k. From Lemma 1 we obtain the following key observation:

Observation 3. *If (L_{k-1}, D_k) is a minimal bad pair, then $\mathbf{S}_k(G) = 0$.*

Now let $\mathscr{A}_k$ be the event that (L_{k-1}, D_k) is a bad $(k-1, k)$ pair. Then

$$Q_k = \Pr(\mathscr{A}_k) \cdot \Pr((L_{k-1}, D_k) \text{ is minimal bad} \mid \mathscr{A}_k) .$$

Observation 3 implies that

$$Q_k \leq \Pr(\mathscr{A}_k) \Pr(\mathbf{S}_k(G) = 0) \mid \mathscr{A}_k) . \tag{2}$$

In the next section, we estimate the right side of (2). Afterwards, these estimates will be put back into (1) to prove that $E(\mathbf{M}) = O(\log n/n)$.

2. APPLICATION OF KNUTH'S THEOREM

To estimate the right side of (2), we need Knuth's generalization of Prüfer codes. Some notation is required to state the result. Suppose $C_1, C_2, \ldots, C_r$ are disjoint, nonempty sets whose union is $[2n]$, and suppose $\mathscr{T}$ is a tree whose r vertices are $C_1, C_2, \ldots, C_r$. Given this "color constraint tree" $\mathscr{T}$, we define another graph $\mathscr{H}_{\mathscr{T}}$ as follows. The vertex set of $\mathscr{H}_{\mathscr{T}}$ is $[2n]$. Suppose that $v \in C_i$ and $w \in C_j$. Then v and w are adjacent in $\mathscr{H}_{\mathscr{T}}$ if and only if C_i and C_j are adjacent in $\mathscr{T}$. The following is a special case of a general theorem of Knuth.

Lemma 2. *(Knuth [2, 4]). The number of spanning trees in $\mathscr{H}_{\mathscr{T}}$ is*

$$\prod_{i=1}^{r} |N_i|^{|C_i|-1} \cdot |C_i|^{\rho_{\mathscr{T}}(C_i)-1} ,$$

where $N_i := \bigcup_{C \in N_{\mathscr{T}}(C_i)} C$, and $\rho_{\mathscr{T}}(C_i)$ is the degree of vertex C_i in $\mathscr{T}$.

Lemma 2 is used several times in this paper. First we use it to give an easy proof of

Lemma 3. *(Palmer).* $Pr(\mathcal{A}_k) = (n^{n-3}(n-k)^{n-k+1}(k-1)^k/n^{2n-2})^2$.

Proof. Let $C_1 = L_{k-1}$, let $C_2 = L_n - L_{k-1} = \{l_k, l_{k+1}, \ldots, l_n\}$, let $C_3 = D_k$, and let $C_4 = D_n - D_k$. Shown in Figure 1 below is a tree $\mathcal{T}$ whose four vertices are labeled C_1, C_2, C_3, and C_4. Observe that

$$\Pr(\mathcal{A}_k) = \Pr(T_1 \text{ and } T_2 \text{ are spanning trees of } \mathcal{H}_{\mathcal{T}}) .$$

By Lemma 2, the number of spanning trees in $\mathcal{H}_{\mathcal{T}}$ is $n^{n-3}(n-k)^{n-k+1}(k-1)^k$. Since the total number of bipartite trees with bipartition (L_n, D_n) is n^{2n-2}, we have

$$\Pr(T_1 \text{ and } T_2 \text{ are spanning trees of } \mathcal{H}_{\mathcal{T}}) = (n^{n-3}(n-k)^{n-k+1}(k-1)^k/n^{2n-2})^2 \quad ■$$

Corollary. *There is an n-independent constant c such that, for all* $k \in [2, \lceil n/2 \rceil]$,

$$\binom{n}{k}\binom{n}{k-1} \Pr(\mathcal{A}_k) < \frac{c}{n} .$$

The second factor in (2) is $\Pr(\mathbf{S}_k = 0 \mid \mathcal{A}_k)$. To estimate it, we use a second moment argument. Observe that

$$E(\mathbf{S}_k \mid \mathcal{A}_k) = \sum_{j=1}^{k-1} \Pr(\rho_{H_k}(l_j) = 0 \mid \mathcal{A}_k) = (k-1) \Pr(\rho_{H_k}(l_{k-1}) = 0 \mid \mathcal{A}_k) . \quad (3)$$

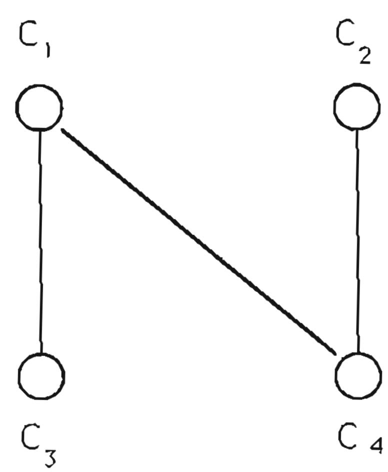

Fig. 1. The color constraint tree.

Putting Lemma 5 and the corollary to Lemma 4 into (6), we get $\Pr(\mathbf{S}_k = 0 \mid \mathcal{A}_k) = O(1/k)$. More precisely, there is an n-independent constant c such that, for all $k \in [2, \lfloor n/2 \rfloor]$,

$$\Pr(\mathbf{S}_k = 0 \mid \mathcal{A}_k) < c/k . \tag{7}$$

Putting (7) and the corollary to Lemma 3 back into (6), we get

$$E(\mathbf{M}) < \sum_{k=2}^{\lfloor n/2 \rfloor} \frac{c}{nk} = O\left(\frac{\log n}{n}\right) \qquad ■$$

Lower Bound. An easy lower bound is presented here for purposes of comparison. Let $\mathbf{Y}(G)$ be the number of bad $(1,2)$ pairs. Then $1 - P_2(n) \ge \Pr(\mathbf{Y} \ne 0)$. Using arguments similar to those in Section 2, one can verify that

$$E(\mathbf{Y}) = \frac{1}{2e^4 n}\left[1 + O\left(\frac{1}{n}\right)\right],$$

and that

$$\begin{aligned} &E(\mathbf{Y}(\mathbf{Y}-1)) \\ &= \binom{n}{2} n \binom{n-2}{2} (n-1) \left(\frac{n^{n-5}(n-4)^{n-1}}{n^{2n-2}}\right)^2 + 3\binom{n}{3}\binom{n}{1}\left(\frac{n^{n-4}(n-3)^{n-1}}{n^{2n-2}}\right)^2 \\ &= \left(\frac{1+2e^2}{4e^8}\right)\frac{1}{n^2}\left[1 + O\left(\frac{1}{n}\right)\right]. \end{aligned}$$

Combining these estimates with the well-known [1] inequality

$$\Pr(\mathbf{Y} = 0) \le 1 - E(\mathbf{Y}) + E(\mathbf{Y}(\mathbf{Y}-1)),$$

we get $\Pr(\mathbf{Y} \ne 0) > c/n$. This proves the following:

Proposition. *There is a constant $c > 0$ such that, for all sufficiently large n,*

$$1 - P_2(n) > c/n .$$

ACKNOWLEDGMENT

I am indebted to Professor Karonski who, after reading an early version of this paper, called my attention to the closely related work of Walker [illegible]

REFERENCES

[1] B. Bollobás, *Random Graphs*, Academic, New York, 1985, [illegible]
[2] D. E. Knuth, Another enumeration of trees, *Can. J. Math.* [illegible]

[3] L. Lovasz and D. Plummer, *Matching Theory*, Annals of Discrete Math **29**, North-Holland, Amsterdam, 1986.

[4] J. W. Moon, Counting labelled trees, *Can. Math. Cong.*, Montreal, 1970.

[5] E. M. Palmer, Matchings in random superpositions of bipartite trees, *J. Comp. Appl. Math* **41**, 199–204 (1992).

[6] H. I. Scoins, On the number of trees with nodes of alternate parity, *Proc. Camb. Phil. Soc.*, **58**, 12–16 (1962).

[7] D. W. Walkup, On the expected value of a random assignment problem, *SIAM J. Comput.* **8**(3), 440–442 (1979).

[8] D. W. Walkup, Matchings in random regular bipartite digraphs, *Disc. Math.*, **31** 59–64 (1980).