This is to certify that the

dissertation entitled

INDEPENDENT SETS IN (r,s)-TREES

presented by

Junghee Cho

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Mathematics

Edgar M. Palmer
Major professor

Date July 15, 1992

INDEPENDENT SETS IN (r, s)-TREES

By

Junghee Cho

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the Degree of

DOCTOR OF PHILOSOPHY

Department of Mathematics

1992

ABSTRACT

INDEPENDENT SETS IN (r, s)-TREES

BY

Junghee Cho

An (r, s)-tree is a connected, acyclic, bipartite graph with r light and s dark vertices. In this thesis, three variable, exponential generating functions are used to find exact values of the expected value $\mu(r, r)$ of the vertex independence number $\beta_0(T)$ of (r, r)-trees T for r up to 19. Also the probabilistic method is applied to find bounds for $\beta_0(T)$ and for the edge independence number $\beta_1(T)$ for almost all (n, n)-trees. These results compare favorably with corresponding bounds for random bipartite graphs.

ACKNOWLEDGMENTS

I would like to express my gratitude to Professor Edgar M. Palmer, my dissertation advisor, for his patience and constant encouragement, and to Professor Bruce Sagan, for his many helpful suggestions. I also thank Professors Jeanne W. Kerr, Wei-Eihn Kuan and Mary J. Winter.

Contents

List of Tables

List of Figures

INTRODUCTION

Graphical enumeration, especially the enumeration of trees is an important topic in graph theory with a long history as well as many applications to various fields, especially chemistry and computer science. Caley first derived a formula for the number of labeled trees [C89]. Since that time the literature has grown substantially. There were 162 references in Moon's definitive monograph *Counting Labelled Trees* [M70] and we suspect that the current number of articles published on this subject is at least 1,000. Included here are papers by Moon and Meir which focus on the independence numbers of trees of various types. In [MeM73] they derived a formula for the expected vertex independence number of a random labeled tree. They were also able to determine the asymptotic value of the expectation. They continued to explore this problem for recursive trees, planted plane trees and tri-valent trees in [MeM75] and found exact and asymptotic formulas for the expectation in each of these cases. This thesis is also concerned with the exact and asymptotic behavior of the independence number of trees.

Here are the few graph theory definitions that we require. A *graph* G consists of a finite nonempty set V of *vertices* and a set E of *edges* which are unordered pairs of distinct vertices. The cardinality of V is called the *order* of G while the cardinality of E is called the *size* of G. An edge e joining the vertices of u and v is denoted by uv and the vertex u and v are said to be *adjacent*. The degree of a vertex v is the number of vertices adjacent to v. A *walk* in a graph G is a sequence of vertices $w_1, w_2, ..., w_m$ such that w_i is adjacent to w_{i+1} for $i = 1$ to $m - 1$. A *path* is a walk in which no vertices are repeated. A *cycle* is a walk with at least 3 different vertices that has no repeated vertices except the first and last. A graph is *connected* if every pair of vertices is joined by a path. A *bipartite* graph is a graph with two kinds of vertices, namely light or dark, and every edge in the graph has one light and one dark vertex. A *tree* is a connected acyclic graph. A *rooted tree* has one vertex, called the *root*, which is distinguished from the others.

A subset S of vertices of a graph is called *independent* if no two vertices of S are adjacent. The *vertex independence number* of a graph G is the number $\beta_0(G)$ of vertices in any largest independent subset of vertices of G. A subset E of edges of a graph is called *independent* if no two edges of E are adjacent. The *edge independence number* of a graph G is the number $\beta_1(G)$ of edges in any largest independent subset of edges of G.

An *(r,s)-tree* is a connected, acyclic, bipartite graph with r light and s dark vertices. Let $\Gamma(r,s)$ denote the set of all (r,s)-trees with r light vertices labeled 1, ..., r and s dark vertices labeled 1, ..., s. We give $\Gamma(r,s)$ the uniform probability distribution. Let $\mu(r,s)$ denote the expected value of $\beta_0(T)$ for (r,s)-trees T in $\Gamma(r,s)$.

Our first objective in this thesis is to determine $\mu(r,s)$ for small values of r and s. The main tool here is an exponential generating function in three variables. Following Moon and Meir [MeM73], we make use of two special types of rooted trees, but our sample space is more complicated because of the two sets of vertices under consideration. Certain operations on these generating functions lead to functional equations from which recurrence relations can be derived. From these, in turn, exact values can be computed.

Secondly, we wish to estimate the asymptotic behavior of $\mu(r,s)$. For this task we use the probabilistic method, which originated in the work of Erdös [Er47], and has been well documented in several books, such as [Bo85], [Pa85] and [AlSE92]. To implement the method, we follow the technique of [Pa92], which employed the matrix-tree theorem to estimate β_1 for random superpositions of (n,n)-trees. We are able to establish bounds on both β_0 and β_1 for almost all (n, n)-trees. Finally these results are compared with a similar treatment of random bipartite graphs.

Chapter 1

ENUMERATION OF INDEPENDENT SETS IN (r,s)-TREES

1.1 *Generating Functions*

Our aim is to determine the expected value of the independence number of (r,s)-trees. As is customary we begin with rooted trees. Let $\Gamma'(r,s)$ be the set of all (r,s)-trees obtained by rooting each tree in $\Gamma(r,s)$ at a vertex. Thus

$$|\Gamma'(r,s)| = |\Gamma(r,s)|(r+s).$$

Let $h_{k,r,s}$ denote the number of rooted trees T in $\Gamma'(r,s)$ such that $\beta_0(T) = k$. The associated exponential generating function (egf) is

$$H = H(x,y,z) = \sum_{r+s\geq 1}^{\infty} (\sum_{k=1}^{r+s} h_{k,r,s} z^k) \frac{x^r}{r!} \frac{y^s}{s!}.$$

Austin[A60] and Scoins [S62] found that the cardinality of the sample space $\Gamma(r,s)$ which is given by

$$|\Gamma(r,s)| = r^{s-1} s^{r-1}.$$

From this it follows that

$$\sum_{k=1}^{r+s} h_{k,r,s} = r^{s-1} s^{r-1}(r+s),$$

which is the number of rooted (r, s)-trees. Hence on setting $z = 1$ in the egf for H we have

$$h(x, y) = H(x, y, 1) = \sum_{r+s\geq 1}^{\infty} r^{s-1}s^{r-1}(r+s)\frac{x^r}{r!}\frac{y^s}{s!}.$$

Ordinarily it is easy to obtain functional relations for generating functions that enumerate rooted trees with special properties. But this case is complicated by the fact that we are dealing not only with several variables but also there are two types of rooted trees with a given independence number. These two types are described in a lemma of Meir and Moon [MeM73] that holds for any rooted tree. Let T denote a tree that is rooted at some vertex v. If every set of $\beta_0(T)$ independent vertices of T contains v we say T is of *type I*. On the other hand, if at least one set of $\beta_0(T)$ independent vertices of T does not contain v, we say T is of *type II*. If we remove the root v of T we obtain a (possibly empty) collection of rooted trees $U_1, ..., U_j$ whose roots were originally joined to v. These are called the branches of T at v and the following lemma relates the independence number of T to the independence number of its branches.

Lemma 1.1.1 *Let T be a type I tree rooted at v. Then each of the rooted trees $U_1, ..., U_j$ is a type II tree, and*

$$\beta_0(T) = 1 + \sum_{i=1}^{j} \beta_0(U_i). \tag{1.1}$$

Proof. Suppose T is a type I tree rooted at v. Let W be a maximum independent set of vertices of T. Since T is a type I tree, v must belong to W. Therefore the roots, $u_1, ..., u_j$ of the trees $U_1, ..., U_j$, respectively, are not in W. We assert that for each $i = 1$ to j, $W \cap V(U_i)$ is a maximum independent set in the branch U_i. It follows that each U_i is a type II tree. Suppose our assertion is false for some i. Then there is a set S of independent vertices in U_i such that

$$|S| > |W \cap V(U_i)|.$$

There are two cases to consider.

(a) $u_i \notin S$. Then we form the set

$$W' = (W \setminus (W \cap V(U_i)) \cup S,$$

of independent vertices in T. Clearly $|W'| > |W|$, a contradiction.

(b) $u_i \in S$. Here we form the set

$$W' = (W \setminus ((W \cap V(U_i)) \cup \{v\})) \cup S,$$

which is independent in T. But $|W'| \geq |W|$ and W' does not include v. So this contradicts the assumption that T is a type I tree and also establishes the formula (1.1) for the vertex independence number of a type I tree. □

The case for type II trees is covered by next lemma whose proof is similar to the one above.

Lemma 1.1.2 *Suppose T is a type II tree rooted at v. Then at least one of the rooted trees $U_1, ..., U_j$ is a type I tree and hence*

$$\beta_0(T) = \sum_{i=1}^{j} \beta_0(U_i). \tag{1.2}$$

Proof. Assume that we have a tree T rooted at v, in which each branch U_i rooted at u_i for $i = 1$ to j is a type II tree. Let W be a maximum independent set of vertices of T and suppose the root v is not in W. But since U_1 is a type II tree, it has an independent set S of vertices, at least as large as $W \cap V(U_1)$ which does not include u_1. Hence we can form the set

$$W' = (W \setminus (W \cap V(U_1))) \cup S,$$

which is an independence set of vertices of T, at least as large as W and does not include u_1. We can treat the other neighbors of v in a similar fashion and arrive at a maximum independent set W'', which does not contain any neighbor of v. Hence $W'' \cup \{v\}$ is a larger independent set than W, a contradiction. □

Now the lemma can be applied to express the exponential generating function $H(x,y,z)$ in terms of generating functions for the two types of trees. Let $g_{k,r,s}$ and $f_{k,r,s}$ denote the numbers of rooted (r,s)-trees with independence number k of type I and type II, respectively. Define the generating functions for the two types by

$$G = G(x,y,z) = \sum_{r+s\geq 1}^{\infty} \left(\sum_{k=1}^{r+s} g_{k,r,s} z^k\right) \frac{x^r}{r!} \frac{y^s}{s!},$$

and

$$F = F(x,y,z) = \sum_{r+s\geq 1}^{\infty} \left(\sum_{k=1}^{r+s} f_{k,r,s} z^k\right) \frac{x^r}{r!} \frac{y^s}{s!}.$$

Obviously

$$h_{k,r,s} = g_{k,r,s} + f_{k,r,s},$$

and $H = G + F$.

We require notation for all of the different kinds of rooted trees. Let H_L and H_D denote the generating functions for (r,s)-trees rooted at a light and a dark vertex respectively. Similarly for G and F, we can define G_L, G_D and F_L, F_D. Then

$$H = H_L + H_D, \quad G = G_L + G_D, \quad F = F_L + F_D$$

and

$$H_L = G_L + F_L, \tag{1.3}$$

$$H_D = G_D + F_D. \tag{1.4}$$

1.2 *Functional Relations*

Now we establish the functional equations that relate these generating functions.

Lemma 1.2.1 *The generating functions G_L, G_D, F_L, F_D, H_L, and H_D for the various types of rooted (r,s)-trees, satisfy the following relations*

$$G_L = zxe^{F_D} \tag{1.5}$$

$$G_D = zye^{F_L} \tag{1.6}$$

$$F_L = x(e^{G_D} - 1)e^{F_D} \tag{1.7}$$

$$F_D = y(e^{G_L} - 1)e^{F_L} \tag{1.8}$$

and

$$H_L = zxe^{F_D} + x(e^{G_D} - 1)e^{F_D} \tag{1.9}$$

$$H_D = zye^{F_L} + y(e^{G_L} - 1)e^{F_L} \tag{1.10}$$

Proof. Suppose T is a type I tree rooted at a light vertex, then by Lemma 1.1.1, each of the rooted branches $U_1, ..., U_j$ obtained by removing the root of T is a type II tree which is rooted at a dark vertex. The generating function for the families of those type II trees, is $F_D^j/j!$ for $j = 0, 1, ...$ Therefore the generating function, G_L for T is;

$$G_L = zx\sum_{j=0}^{\infty} F_D^j/j! = zxe^{F_D}.$$

The factor x is present to account for the root of T and the factor z is present because of equation (1.1) of Lemma 1.1.1.

Suppose T is a type II tree rooted at a light vertex, then by Lemma 1.1.2, at least one of the rooted branches $U_1, ..., U_j$ must be a type I tree which is rooted at a dark vertex. Therefore the generating function F_L for T has a factor of

$$\sum_{j=1}^{\infty} G_D^j/j! = e^{G_D} - 1,$$

the generating function for non-empty families of type I trees . There may or may not be some type II trees among the branches $U_1, ..., U_j$ and the generating function for (possibly empty) families of type II trees rooted at dark vertices is e^{F_D}, as before. The observations imply that

$$F_L = x(e^{G_D} - 1)e^{F_D}.$$

The factor x is present to account for the root of T, but because of equation (1.2) of Lemma 1.1.2, the factor z is not included here. Equations (1.6) and (1.8) follow from (1.5) and (1.7) respectively by symmetry and equations (1.9) and (1.10) follow from (1.5) through (1.8) combined with (1.3) and (1.4). □

The formulas in this lemma can be used to calculate both the number of rooted trees of the various kinds with $r + s$ vertices as well as the sum of the independence numbers of these trees. To handle the former task, we set $z = 1$ in the various egf's introduced earlier and we also simplify the notation as follows. Let

$$g = g(x, y) = G(x, y, 1), \quad f = f(x, y) = F(x, y, 1), \quad h = h(x, y) = H(x, y, 1)$$

and similarly

$$g_L = g_L(x, y) = G_L(x, y, 1), \quad g_D = g_D(x, y) = G_D(x, y, 1),$$

with the same notation for f_L, f_D and h_L, h_D.

Furthermore, it follows from (1.9), (1.10) and the equation $H = H_L + H_D$ that

$$h = xe^{h_D} + ye^{h_L}.$$

This completes the treatment of the relevant functional equations for the several types of rooted trees.

1.3 *Recurrence Relations and Numerical Values*

First we introduce notation for the coefficients of the simplified egf's of the previous section:

$$g_L = \sum_{r+s\geq 1}^{\infty} g_{L(r,s)} \frac{x^r}{r!}\frac{y^s}{s!}, \quad g_D = \sum_{r+s\geq 1}^{\infty} g_{D(r,s)} \frac{x^r}{r!}\frac{y^s}{s!},$$

$$f_L = \sum_{r+s\geq 1}^{\infty} f_{L(r,s)} \frac{x^r}{r!}\frac{y^s}{s!}, \quad f_D = \sum_{r+s\geq 1}^{\infty} f_{D(r,s)} \frac{x^r}{r!}\frac{y^s}{s!},$$

$$h_L = \sum_{r+s\geq 1}^{\infty} h_{L(r,s)} \frac{x^r}{r!}\frac{y^s}{s!}, \quad h_D = \sum_{r+s\geq 1}^{\infty} h_{D(r,s)} \frac{x^r}{r!}\frac{y^s}{s!},$$

so that

$$\sum_{k=1}^{r+s} h_{k,r,s} = h_{L(r,s)} + h_{D(r,s)}.$$

The following lemma provides recurrence relations for the coefficients of the generating functions above .

Lemma 1.3.1 *Let r and s be non negative integers. Then*

$$g_{L(1,0)} = 1, \tag{1.11}$$

$$g_{L(r,s)} = 0, \text{ for } r = 0,1 \text{ and } s > 0 \tag{1.12}$$

and for $r > 1$,

$$g_{L(r,s)} = \sum \frac{r-i}{r-1} \binom{r}{i} \binom{s}{j} g_{L(i,j)} f_{D(r-i,s-j)}, \tag{1.13}$$

where the sum runs over all i, j such that $0 < i < r$ and $0 \leq j \leq s$.

Also

$$f_{L(1,0)} = 0 \text{ and } f_{L(r,s)} = r \text{ for } r = 0,1 \text{ and } s > 0, \tag{1.14}$$

and for $r > 1$,

$$f_{L(r,s)} = \sum \frac{r-i}{r-1} \begin{pmatrix} r \\ i \end{pmatrix} \begin{pmatrix} s \\ j \end{pmatrix} [f_{L(i,j)}(g_{D(r-i,s-j)} + f_{D(r-i,s-j)}) + g_{L(i,j)} g_{D(r-i,s-j)}], \tag{1.15}$$

where the sum runs over all i, j *such that* $0 < i < r$ *and* $0 \leq j \leq s$.

Proof. The initial values in (1.11) and (1.12) are easily found by considering the tree diagrams. On setting $z = 1$ in equation (1.5), we find

$$g_L = xe^{f_D}, \tag{1.16}$$

and on differentiating this equation with respect to x we have

$$\frac{\partial g_L}{\partial x} = e^{f_D} + xe^{f_D} \frac{\partial f_D}{\partial x}. \tag{1.17}$$

On solving (1.16) for e^{f_D} and substituting the result in (1.17), the latter becomes

$$x \frac{\partial g_L}{\partial x} = g_L(1 + x \frac{\partial f_D}{\partial x}). \tag{1.18}$$

Finally, using the definitions of g_L and f_D, we can compare coefficients on both sides of (1.18) to obtain (1.13).

Again, the initial values in (1.14) can be found easily from the tree diagrams. Next we verify (1.15). On setting $z = 1$ in equation (1.7), we find

$$f_L = x(e^{g_D} - 1)e^{f_D}, \tag{1.19}$$

and on differentiating this equation with respect to x we obtain

$$\frac{\partial f_L}{\partial x} = (e^{g_D} - 1)e^{f_D} + xe^{f_D} e^{g_D} \frac{\partial g_D}{\partial x} + x(e^{g_D} - 1)e^{f_D} \frac{\partial f_D}{\partial x}, \tag{1.20}$$

which leads to

$$x\frac{\partial f_L}{\partial x} - f_L = f_L x(\frac{\partial g_D}{\partial x} + \frac{\partial f_D}{\partial x}) + g_L x\frac{\partial g_D}{\partial x}. \tag{1.21}$$

By comparing coeffients of both sides of (1.21) we obtain (1.15). □

There is also a combinatorial proof of this theorem that was observed by B.Sagan, involving an extra root. For example to prove (1.13), consider a type I, (r,s)-tree T with a root and an extra root w both in light vertices. Let U be the branch of T which contains w. Next we deform T into two parts, $T \setminus U$ and U, then identify them as follows:

$$T \longleftrightarrow (T \setminus U, U). \tag{1.22}$$

For each fixed extra root of T, $g_{L(r,s)}$ is the number of trees in the left side of (1.22) and there are $(r-1)$ ways to label the extra root. Therefore for $r > 1$, $(r-1)g_{L(r,s)}$ is the number of trees T in the left side of (1.22). On the other hand, $T \setminus U$ is a type I tree by Lemma 1.1.1 and counted by $g_{L(i,j)}$ for some i, j up to labels and U is a type II tree counted by $(r-i)f_{D(r-i,s-j)}$ up to labels. Of course $\begin{pmatrix} r \\ i \end{pmatrix}\begin{pmatrix} s \\ j \end{pmatrix}$ accounts for labeling. Hence the right side of (1.22) is counted by

$$\sum_{\substack{0<i<r,\\ 0\leq j\leq s}} \begin{pmatrix} r \\ i \end{pmatrix}\begin{pmatrix} s \\ j \end{pmatrix} g_{L(i,j)}(r-i)f_{D(r-i,s-j)}. \tag{1.23}$$

And (1.13) follows.

The same correspondence as (1.22) can be used to show (1.15). Let T be a type II, (r,s)-tree with a root and an extra root w both in light vertices. Then for $r > 1$, $(r-1)f_{L(r,s)}$ is the number of trees T in the left side of (1.22). But this time $T \setminus U$ can be either type I or type II. If $T \setminus U$ is a type I tree, then U must be a type I to make T a type II. And $(r-i)g_{L(i,j)}g_{D(r-i,s-j)}$ will count those trees up to labels. On the other hand if $T \setminus U$ is a type II tree, then U can be either a type I or a type

II. And $(r-i)\{f_{L(i,j)}(g_{D(r-i,s-j)}+f_{D(r-i,s-j)})\}$ will be the number of these trees up to labels. Hence (1.15) follows.

Observe that Lemma 1.3.1 serves to calculate values of $g_{D(r,s)}$ and $f_{D(r,s)}$ because of the symmetric relation between light and dark. Specifically

$$g_{D(r,s)} = g_{L(s,r)} \quad \text{and} \quad f_{D(r,s)} = f_{L(s,r)}.$$

Now we can find the number of trees of different types for small values of r and s. The computations begin with the boundary values in (1.11), (1.12) and (1.14). We have used the lemma to count trees for all r and $s \leq 19$. We include in Table 1.1 the data for $r+s=9$ and $r+s=10$.

The last column, $\sum_{k=1}^{r+s} h_{k,r,s}$, is obtained by summing all prior columns, i.e.

$$\sum_{k=1}^{r+s} h_{k,r,s} = g_{L(r,s)} + g_{D(r,s)} + f_{L(r,s)} + f_{D(r,s)}.$$

Notice the symmetry of the data in this table. For example, $g_{L(7,2)}$, which is the number of (r,s)-trees of type I with 7 light and 2 dark vertices rooted at a light vertex, is 3122, which is the same as $g_{D(2,7)}$. We can also illustrate the numbers for $(7,2)$-trees with tree diagrams. Consider a $(7,2)$-tree T. We can see easily each maximum independent vertex set contains either none or exactly one dark vertex and its cardinality (the vertex independence number) is 7 for all $(7,2)$-trees. Also there exists exactly one light vertex adjacent to both dark ones for all those trees. Next, besides the unique light vertex which we just mentioned, we have

$$\sum_{i=0}^{6} \binom{6}{i} = 2^6$$

many ways to partition the remaining 6 light vertices into two sets where the vertices of each set are joined to one of the dark vertices as in Figure 1.1. The letters inside the circles indicate the number of light vertices in each set.

Table 1.1: Number of trees of different types

r	s	$g_{L(r,s)}$	$g_{D(r,s)}$	$f_{L(r,s)}$	$f_{D(r,s)}$	$\sum_{k=1}^{r+s} h_{k,r,s}$
8	1	8	0	0	1	9
7	2	3122	0	14	896	4032
6	3	48690	36	3798	26208	78732
5	4	97640	6180	62360	121820	288000
4	5	6180	97640	121820	62360	288000
3	6	36	48690	26208	3798	78732
2	7	0	3122	896	14	4032
1	8	0	8	1	0	9
9	1	9	0	0	1	10
8	2	8176	0	16	2048	10240
7	3	240156	42	9891	107121	357210
6	4	1044120	15912	282984	868824	2211840
5	5	382600	382600	1570525	1570525	3906250
4	6	15912	1044120	868824	282984	2211840
3	7	42	240156	107121	9891	357210
2	8	0	8176	2048	16	10240
1	9	0	9	1	0	10

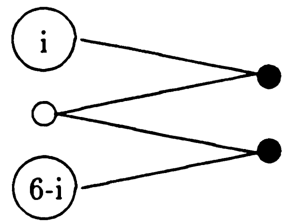


Figure 1.1: A diagram for (7,2)-trees

Since there are 7 ways to label the unique light vertex and 2^6 ways to partition the remaining 6 vertices, there are 7×2^6 labeled $(7,2)$-trees. Hence there are $7 \times 7 \times 2^6$ $(7,2)$-trees rooted at a light vertex, i.e.

$$g_{L(7,2)} + f_{L(7,2)} = 7 \times 7 \times 2^6 = 3136.$$

Next we consider how many of them will be of type II which are counted by $f_{L(7,2)}$. Since they are the ones containing one dark vertex in their maximum independent vertex set, we can easily see them as the case of $i = 0$ or 6 in the above diagram. Hence we have

$$f_{L(7,2)} = 7 \times 2$$

many of them, where 7 is the number of ways to label the unique light vertex and 2 is the number of ways to label the dark vertices. Therefore

$$g_{L(7,2)} = 3136 - f_{L(7,2)} = 3122.$$

Next

$$g_{D(7,2)} = 0$$

is immediate and similarly we can figure

$$f_{D(7,2)} = 7 \times 2^6 \times 2 = 896,$$

where the last factor 2 is the number of ways to root the tree at a dark vertex. And by adding those numbers we obtain the last column 4032.

1.4 *Expectation of the Vertex Independence Number*

As defined in the introduction, $\mu(r,s)$ is the expected value of the vertex independence number with uniform probability distribution assigned to $\Gamma(r,s)$. The computation of $\mu(3,3)$ is illustrated in Table 1.2. The four isomorphism types of (3,3)-trees T are shown together with the number of ways, $l(T)$, to label each one. The vertex independence number $\beta_0(T)$ is also provided. Hence we have

$$\mu(3,3) = (9 \times 4 + 36 \times 3 + 18 \times 3 + 18 \times 3)/81.$$

The table also displays the ratio of the expected value to the order, e.g. $\mu(r,r)/2$r. Of course, this ratio must be at least .5, and our computational results will indicate that it is not ever much bigger than $.56\cdots$.

Our aim is to develop recurrence relations for $\mu(r,s)$. Since the independence number of a tree T is constant for all rooted versions of T, we can express $\mu(r,s)$ in terms of rooted trees as follows:

$$\mu(r,s) = \sum \beta_0(T)/(r^{s-1}s^{r-1}(r+s)),$$

where the sum is over all rooted (r,s)-trees T. Hence

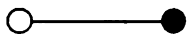
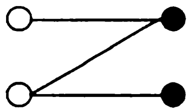
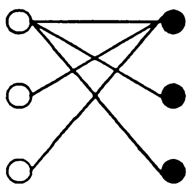
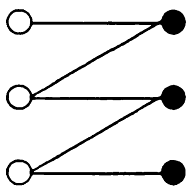
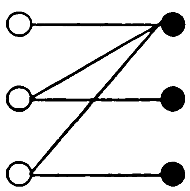
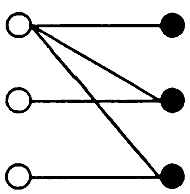
$$\mu(r,s)r^{s-1}s^{r-1}(r+s) = \sum \beta_0(T). \tag{1.24}$$

This motivates the choice of the generating function $M(x,y)$ which delivers the sum of the independence numbers for rooted (r,s)-trees. We define

$$M(x,y) = \sum_{r+s\geq 1}^{\infty} \mu(r,s)r^{s-1}s^{r-1}(r+s)\frac{x^r}{r!}\frac{y^s}{s!}. \tag{1.25}$$

Now our task is to express $M(x,y)$ in terms of the generating functions of the previous sections. The result is given in the following theorem.

Table 1.2: Expectation of vertex independence number

T	$l(T)$	$\beta_0(T)$	$\mu(\mathbf{r},\mathbf{r})$	$\mu(\mathbf{r},\mathbf{r})/\mathbf{2r}$
	1	1	1	.5
	4	2	2	.5
	9	4		
	36	3	252/81=3.111111	252/486=.518518
	18	3		
	18	3		

Theorem 1.4.1 *The generating function M(x,y) for the independence numbers of rooted (r,s)-trees is*

$$M(x,y) = \frac{(1+h_D)g_L + (1+h_L)g_D}{1-h_L h_D}. \tag{1.26}$$

Proof. It follows from our definitions that

$$\sum \beta_0(T) = \sum_k k h_{k,r,s},$$

where the sum on the left is over all rooted (r,s)-trees. Hence

$$\mu(r,s) r^{s-1} s^{r-1}(r+s) = \sum_k k h_{k,r,s}.$$

Therefore

$$M(x,y) = (\partial H/\partial z)_{z=1}.$$

If we differentiate both sides of equation (1.9) with respect to z and simplify using (1.4), then we have

$$\partial H_L/\partial z = xe^{F_D} + xe^{H_D}\partial H_D/\partial z + (z-1)xe^{F_D}\partial F_D/\partial z. \tag{1.27}$$

The companion equation for trees rooted at a dark vertex is obtained from (1.24) by substituting D for L and y for x;

$$\partial H_D/\partial z = ye^{F_L} + ye^{H_L}\partial H_L/\partial z + (z-1)ye^{F_L}\partial F_L/\partial z. \tag{1.28}$$

Next we set $z = 1$ in both equations (1.24) and (1.25) . Recall that we obtained f_L from $F_L(x,y,z)$ by setting $z = 1$, i.e.:

$$f_L = f_L(x,y) = F_L(x,y,1),$$

and

$$f_D = f_D(x,y) = F_D(x,y,1), etc.$$

Then we arrive at the following system of equations for the two partials $(\partial H_L/\partial z)_{z=1}$ and $(\partial H_D/\partial z)_{z=1}$:

$$\begin{cases} (\partial H_L/\partial z)_{z=1} = xe^{f_D} + xe^{h_D}(\partial H_D/\partial z)_{z=1} \\ (\partial H_D/\partial z)_{z=1} = ye^{f_L} + ye^{h_L}(\partial H_L/\partial z)_{z=1} \end{cases}$$

After solving the system, we use the equation

$$H = H_L + H_D$$

to find

$$(\partial H/\partial z)_{z=1} = \frac{xe^{f_D} + ye^{f_L} + xy(e^{h_D+f_L} + e^{h_L+f_D})}{1 - xye^h}. \tag{1.29}$$

The following equations are obtained from (1.5) through (1.8) by setting $z = 1$:

$$g_L = xe^{f_D}$$

$$g_D = ye^{f_L}$$

$$f_L = xe^{g_D+f_D} - xe^{f_D}$$

$$f_D = ye^{g_L+f_L} - ye^{f_L},$$

Then the result of substitution in the right side of (1.26) is

$$(\partial H/\partial z)_{z=1} = \frac{(1+h_D)g_L + (1+h_L)g_D}{1 - h_L h_D},$$

which is the required formula. □

Equation (1.23) in Theorem 1.4.1. can be written as

$$M(x,y) - M(x,y)h_L h_D = g_L + g_D + h_D g_L + h_L g_D, \tag{1.30}$$

which leads to

$$M(x,y) = g_L + g_D + h_D g_L + h_L g_D + M(x,y)h_L h_D. \tag{1.31}$$

By comparing the coefficients of both sides in (1.28) we have the recursive relation of the following corollary.

Corollary 1.4.1 *The expected value of the vertex independence number for the labeled (r,s)-trees, $\mu(r,s)$, has the following recurrence relation:*

$$\mu(r,s) = \frac{1}{r^{s-1}s^{r-1}(r+s)}[g_{L(r,s)}+g_{D(r,s)}+\sum_{\substack{i\leq r\\ j\leq s}} \binom{r}{i}\binom{s}{j}(h_{D(i,j)}g_{L(r-i,s-j)}+h_{L(i,j)}g_{D(r-i,s-j)})$$

$$+\sum_{\substack{i+i'+i''=r\\ j+j'+j''=s}} \binom{r}{i,\ i',\ i''}\binom{s}{j,\ j',\ j''} h_{L(i',i'')}h_{D(j',j'')}\mu(i,j)i^{j-1}j^{i-1}(i+j)],$$

where in the second sum, we have $i < r$ and $j < s$.

Corollary 1.4.1 was used to compute the expected value $\mu(r,r)$ for $r = 1$ to 19. The ratio $\mu(r,r)/2r$ of the expected value to the order of these trees is displayed in Table 1.3. Note that the values for $r = 1$, 2 and 3 agree with the entries in Table 1.2 which were found by inspecting the tree diagrams. We also verified the value of $\mu(4,4)/8$ in Table 1.3 by inspecting all the relevant tree diagrams. Notice also that the sequence of the ratios increases slowly and $\mu(19,19)/38$ is approximately .5615.

Table 1.3: Expected independence number by proportion when $r = s$

r	$\mu(r,r)/2r$
1	0.50000000000000000000
2	0.50000000000000000000
3	0.51851851851851851852
4	0.53173828125000000000
5	0.54026240000000000000
6	0.54582540493382614439
7	0.54959402326760519619
8	0.55225856471315637464
9	0.55422138607749021473
10	0.55572092017749310500
11	0.55690257221809530113
12	0.55785805356475197043
13	0.55864727324199846080
14	0.55931072818637855904
15	0.55987668338524781944
16	0.56036544760401056140
17	0.56079199438507640447
18	0.56116761656041338555
19	0.56150100107406312305

Table 1.4: Total of the vertex independence numbers

$r+s$	r	s	$\sum_{T\in\Gamma(r,s)} \beta_0(T)$
4	1	3	3
	2	2	8
8	1	7	7
	2	6	1152
	3	5	10140
	4	4	17424
12	1	11	11
	2	10	51200
	3	9	4782996
	4	8	67159318
	5	7	265688360
	6	6	396047700
16	1	15	15
	2	14	1605632
	3	13	1167575916
	4	12	86974860336
	5	11	1573585770680
	6	10	10124304409740
	7	9	28269280359936
	8	8	38861741660224

Chapter 2

ASYMPTOTIC BEHAVIOR

2.1 *The probabilistic method*

Our aim in this chapter is to study the asymptotic behavior of the expected vertex independence number of (n, n)-trees. The difficulty of this type of problem has been noted before. For example, Bender ([Be74], p.512) has observed:

"Practically nothing is known about asymptotics for recursions in two variables even when a generating function is available. Techniques for obtaining asymptotics from bivariate generating functions would be quite useful. Some results have been obtained for a small class of generating functions (Bender [2]), but these are often hard to apply."

Nevertheless, we are able to establish bounds on the independence number for most (n, n)-trees. The tools we use include the matrix-tree theorem and the probabilistic method. The matrix-tree theorem originated in the work of Kirkhoff (see Moon [M70] p.42) and relates the number of spanning trees of a labeled graph to the adjacency matrix. In particular, if G is a labeled graph of order n with vertex set

$$V = \{v_1, ..., v_n\},$$

then the adjacency matrix $A(G)$ is the n by n matrix whose i, j entry is 1 or 0 according as vertices v_i and v_j are adjacent or not. The matrix $D(G)$ has diagonal deg v_1, ..., deg v_n, while the off-diagonal entries are all zero.

Theorem 2.1.1 (Matrix-Tree Theorem) *The number of spanning trees in any labeled graph G is the value of any cofactor of $D(G)-A(G)$.*

This theorem has been applied to many important classes of graphs for which convenient formulas have been found. For example, when applied to the complete graph K_n, it yields Cayley's formula, n^{n-2}, for the number of labeled trees [C89]. We will make use of the theorem for a particularly simple but important family. Consider a graph, denoted $G(V_1, V_2, V_3, V_4)$, whose vertex set V is partitioned into four non empty sets:

$$V = V_1 \,\dot{\cup}\, V_2 \,\dot{\cup}\, V_3 \,\dot{\cup}\, V_4.$$

The edge set of $G(V_1, V_2, V_3, V_4)$ consists of all edges joining vertices of V_1 to vertices of V_2, vertices of V_2 to vertices of V_3, and vertices of V_3 to vertices of V_4.

Corollary 2.1.1 *The number of spanning trees of $G(V_1, V_2, V_3, V_4)$ is*

$$(n_1 + n_3)^{n_2 - 1}(n_2 + n_4)^{n_3 - 1} n_2^{n_1} n_3^{n_4},$$

where $n_i = |V_i|$ for $i = 1$ to 4.

This formula can be established in several ways. We obtained it by applying row and column operations to calculate a cofactor of the required matrix. Our colleague, Greg Buzzard, verified these calculations using *Mathematica*. It can also be realized as a corollary of a very broad result of Knuth [Kn68] on generalized Prüfer codes (see also Moon [M70] p.10).

The probabilistic method was first applied to graphs by Erdös [Er47], who pioneered its use with so many innovations that it may be more proper to call it *the*

Erdős method. Here we sketch only the portion of the method that we require. For background on probability theory for a discrete sample space, one can consult the appendix of [Pa85].

Theorem 2.1.2 (Markov's Inequality) *Let $X \geq 0$ be a random variable and let $t > 0$. Then*

$$P(X \geq t) \leq \frac{E(X)}{t}. \tag{2.1}$$

On setting $t = 1$ on inequality (2.1), we have

$$P(X \geq 1) \leq E(X). \tag{2.2}$$

If X is non negative and integer valued, we also have

$$P(X = 0) + P(X \geq 1) = 1. \tag{2.3}$$

In our applications, the sample space always consists of graphs with at least n vertices and the random variable X counts certain types of subgraphs. It follows from (2.2) and (2.3) that if

$$E(X) \to 0 \text{ as } n \to \infty,$$

then

$$P(X \geq 1) \to 0$$

and

$$P(X = 0) \to 1$$

and we say "almost all graphs have no such subgraph". Suppose the sample space consists of (n, n)-trees and X counts sets of vertices which are undesirable, i.e. "bad sets". If we show $E(X) \to 0$, then we say "almost all trees have no bad sets".

2.2 *Vertex independence number for almost all (n,n)-trees*

Our aim is to determine bounds for β_0 for almost all (n,n)-trees in $\Gamma(n,n)$. Of course, we always have the lower bound

$$n \leq \beta_0,$$

for all trees in $\Gamma(n,n)$. To find an upper bound we require some notation and a few more definitions.

Fix $k > n$ and let X_k be the random variable on $\Gamma(n,n)$ which counts the number of independent vertex sets of order k. Each such k-set consists of i light and j dark vertices with

$$0 < i, j < n \text{ and } i + j = k. \tag{2.4}$$

Now let A_k be the set of all trees in $\Gamma(n,n)$ for which a specified set of i light and j dark vertices are independent. Then $|A_k|$ is the number of spanning trees of $G(V_1, V_2, V_3, V_4)$ in Corollary 2.1.1, with $|V_1| = i$, $|V_2| = n - j$, $|V_3| = n - i$ and $|V_4| = j$. By the Corollary ,

$$|A_k| = n^{n-j-1} n^{n-i-1} (n-j)^i (n-i)^j. \tag{2.5}$$

Since $\Gamma(n,n)$ is a sample space with the uniform probability distribution, $P(A_k)$ is the ratio of $|A_k|$ to $|\Gamma(n,n)|$. Hence

$$P(A_k) = n^{n-j-1} n^{n-i-1} (n-j)^i (n-i)^j / n^{2n-2} \tag{2.6}$$

and

$$E(X_k) = \sum_{i+j=k} \binom{n}{i} \binom{n}{j} P(A_k), \tag{2.7}$$

where $\binom{n}{i}$ is the number of ways of choosing i light vertices and $\binom{n}{j}$ is the number of ways of choosing j dark ones.

Next we want to determine k as a function of n, so that,

$$E(X_k) \to 0 \text{ as } n \to \infty. \tag{2.8}$$

Then we can say "almost all (n, n)-trees have no independent sets of order k". This idea is also expressed by saying *almost all* (n, n)*-trees* T *have*

$$\beta_0(T) < k. \tag{2.9}$$

Naturally the function just mentioned should be as small as possible. We set

$$k = (1 + \alpha)n, \tag{2.10}$$

where α is a positive constant, and we will try to determine α as small as possible so that (2.8) holds. Note that the right side of (2.10) is not necessarily an integer. We often use non-integral quantities where we ought to round up or down. It should be clear that such deviations do not affect the validity of the results.

Since $\Gamma(n, n)$ is a sample space with the uniform probability distribution, $P(A_k)$ is the ratio of $|A_k|$ to $|\Gamma(n, n)|$. Hence

$$P(A_k) = n^{n-j-1} n^{n-i-1} (n - j)^i (n - i)^j / n^{2n-2} \tag{2.11}$$

and

$$E(X_k) = \sum_{i+j=k} \binom{n}{i} \binom{n}{j} P(A_k), \tag{2.12}$$

where $\binom{n}{i}$ is the number of ways of choosing i light vertices and $\binom{n}{j}$ is the number of ways of choosing j dark ones.

The conditions above on i and j imply that

$$\alpha n < i, j < n. \tag{2.13}$$

Also because of symmetry in i and j in the formula (2.11) for $P(A_k)$ and in (2.12) for $E(X_k)$, we have the following upper bound for $E(X_k)$.

$$E(X_k) \leq 2\sum \binom{n}{i}\binom{n}{j} P(A_k), \tag{2.14}$$

where sum is for all i, j with

$$\alpha n < i < (n + \alpha n + 1)/2 \text{ and } i + j = (1+\alpha)n. \tag{2.15}$$

Next we simplify (2.14) using Stirling's formula for $n!$. We make a slight change in notation by setting

$$i = t \text{ and } j = (1+\alpha)n - t \tag{2.16}$$

and the result is

$$E(X_k) = O(1)n^{n-\alpha n+1} \sum_{\alpha n<t<(n+\alpha n+1)/2} a_t, \tag{2.17}$$

where

$$a_t = (n-t)^{\alpha n-1/2}(t-\alpha n)^{\alpha n-1/2}/(n+\alpha n-t)^{n+\alpha n-t+1/2}t^{t+1/2}. \tag{2.18}$$

To see the behavior of the series, we investigate the ratio, a_{t+1}/a_t:

$$a_{t+1}/a_t = f_1^{\alpha n-1/2} f_2 f_3^{n-t} f_4^{t-\alpha n+1}, \tag{2.19}$$

where

$$f_1 = \frac{(n-t-1)(t-\alpha n+1)(n+\alpha n-t)t}{(n-t-1)(t-\alpha n+1)(n+\alpha n-t)t + \alpha n^2(2t-n-\alpha n+1)}, \tag{2.20}$$

$$f_2 = (n+\alpha n-t)/(t+1), \tag{2.21}$$

$$f_3 = (n+\alpha n-t)/(n+\alpha n-t-1) \tag{2.22}$$

and

$$f_4 = t/(t+1). \tag{2.23}$$

Notice all the exponents of the f_i are non negative for sufficiently large n and it is obvious that

$$f_1 \text{ and } f_2 > 1 \text{ if } t < (n + \alpha n - 1)/2. \tag{2.24}$$

Next we consider $f_3^{n-t} f_4^{t-\alpha n+1}$ of (2.19):

$$f_3^{n-t} f_4^{t-\alpha n+1} = (\frac{n + \alpha n - t}{n + \alpha n - t - 1})^{n-t} (\frac{t}{t+1})^{t-\alpha n+1} \tag{2.25}$$

$$= (1 + \frac{1}{n + \alpha n - t - 1})^{n+\alpha n-t-1} (1 + \frac{1}{t})^{-t} (1 + \frac{n + \alpha n - 1 - 2t}{(n + \alpha n - t)t})^{\alpha n-1}. \tag{2.26}$$

Observe that

$$(1 + \frac{1}{n + \alpha n - t - 1})^{n+\alpha n-t-1} (1 + \frac{1}{t})^{-t} > 1 \text{ if } n + \alpha n - t - 1 > t. \tag{2.27}$$

This follows from the fact that

$$f(x) = (1 + \frac{1}{x})^x \tag{2.28}$$

increases to e as $x \rightarrow \infty$. It is also clear that for the last factor of (2.26), we have

$$(1 + \frac{n + \alpha n - 1 - 2t}{(n + \alpha n - t)t})^{\alpha n-1} > 1 \text{ if } t < (n + \alpha n - 1)/2. \tag{2.29}$$

Hence we find that

$$a_{t+1}/a_t > 1 \text{ if } t < (n + \alpha n - 1)/2, \tag{2.30}$$

i.e. the series increases all the way to the very last term, which is a_{t_0}, where

$$t_0 = \lfloor (n + \alpha n + 1)/2 \rfloor. \tag{2.31}$$

Now the sum in (2.14) is bounded by the product of the last term and the length of the sum. Hence we have

$$E(X_k) = O(1) n^{n-\alpha n+1} (n - \alpha n - 1) a_{t_0}, \tag{2.32}$$

where the factor $(n - \alpha n - 1)$ is contributed by the length of the sum. On the other hand,

$$a_{t_0} = O(1)a_{(n+\alpha n)/2} \tag{2.33}$$

and a bit of algebra shows that

$$a_{(n+\alpha n)/2} = O(1)n^{-(n-\alpha n+2)}(2^{1-\alpha}(1-\alpha)^{2\alpha}/(1+\alpha)^{1+\alpha})^n. \tag{2.34}$$

Substituting (2.33) and (2.34) in the equation (2.32), we find that the upper bound for the expectation takes the following simple form:

$$E(X_k) = O(1)(2^{1-\alpha}(1-\alpha)^{2\alpha}/(1+\alpha)^{1+\alpha})^n. \tag{2.35}$$

Since we want the right side of (2.35) to approach zero as $n \to \infty$, we just need to solve the inequality

$$2^{1-\alpha}(1-\alpha)^{2\alpha} < (1+\alpha)^{1+\alpha}. \tag{2.36}$$

A simple numerical calculation shows that it is sufficient to choose

$$\alpha = .27974. \tag{2.37}$$

These observations are summerized in the following theorem.

Theorem 2.2.1 *For almost all (n,n)-trees T, the vertex independence number $\beta_0(T)$ satisfies the inequality:*

$$n \le \beta_0(T) \le (1.27974)n. \tag{2.38}$$

We now show that the upper bound in (2.38) cannot be significantly improved using the probabilistic method. To do this we will show that

$$E(X_k) \to \infty \text{ as } n \to \infty,$$

for a slightly smaller value of $k = (1+\alpha)n$. An argument similar to the one which produced (2.17) also shows that for some constant $c_0 > 0$,

$$E(X_k) \geq c_0 n^{n-\alpha n+1} \sum_{\alpha n < t < (n+\alpha n+1)/2} a_t, \tag{2.39}$$

where a_t is already defined in (2.18). We have shown that the terms in the sum on the right side of (2.39) are increasing. Therefore the sum is bounded below by

$$(t_0 - t_1) a_{t_1}, \tag{2.40}$$

for any t_1 in the interval of summation. It is convenient to describe t_1 as follows:

$$t_1 = \lfloor (1-\varepsilon)t_0 + \varepsilon \alpha n \rfloor, \text{ where } 0 < \varepsilon < 1. \tag{2.41}$$

Then the lower bound in (2.39) becomes

$$E(X_k) \geq c_0 n^{n-\alpha n+1} \frac{\varepsilon(1-\alpha)}{2} n a_{t_1}. \tag{2.42}$$

Now the right side of (2.42) can be evaluated using the formula for a_t. After a bit of work we find that

$$E(X_k) \geq c_1 B^n, \tag{2.43}$$

where c_1 is a positive constant and

$$B = \frac{2^{(1-\alpha)}(1-\alpha)^{2\alpha}(1+\varepsilon)^{\alpha}(1-\varepsilon)^{\alpha}}{\sqrt{(1+\alpha+\varepsilon(1-\alpha))^{1+\alpha+\varepsilon(1-\alpha)}}\sqrt{(1+\alpha-\varepsilon(1-\alpha))^{1+\alpha-\varepsilon(1-\alpha)}}}. \tag{2.44}$$

A numerical calculation shows that if we drop the value of α down a bit from .27974 then B will be bigger than 1 for certain values of ε. For example if we take

$$\alpha = .2797 \text{ and } \varepsilon = .001,$$

we find

$$B = 1.000134\cdots. \tag{2.45}$$

Therefore

$$E(X_k) \to \infty \text{ as } n \to \infty \text{ for } k = (1.2797)n. \tag{2.46}$$

This shows that the upper bound in the theorem cannot be reduced substantially.

Our exact calculations in Chapter 1 indicate that the expected value of β_0 is about

$$(.5615\cdots)2n = (1.1230\cdots)n.$$

The latter number is in the middle of the interval described in Theorem 2.2.1. We suspect that the value of β_0 for almost all (n,n)-trees is even more closely concentrated about the asymptotic value of the mean than the interval of Theorem 2.2.1.

2.3 *Edge independence number for almost all (n, n)-trees*

Recall that the edge independence number of a tree T is denoted by $\beta_1(T)$. Our aim in this section is to find bounds for $\beta_1(T)$ for almost all (n,n)-trees T in $\Gamma(n,n)$. This was a problem left unsolved in [Pa92] and [Sch92]. The following theorem shows the close relationship of β_0 and β_1 for trees.

Theorem 2.3.1 *For any tree T of order n,*

$$\beta_0(T) + \beta_1(T) = n. \tag{2.47}$$

This result follows quickly from theorems of Gallai[Ga59] and König[Kö31] (see also [H69] pp.95-96). It can also be derived directly by an algorithmic approach that produces the required independent sets.

Now this theorem and Theorem 2.2.1 can be combined to provide bounds for β_1 which we state in the next theorem.

Theorem 2.3.2 *For almost all (n,n)-trees T, the edge independence number* $\beta_1(T)$ *satisfies the inequality:*

$$(.72026)n \leq \beta_1(T) \leq n. \tag{2.48}$$

There is another approach which could lead to an improvement in the lower bound in (2.48). It stems from an idea in [Pa92] for finding matchings in superpositions of trees. This alternative makes use of the following slight generalization of Hall's theorem [Ha35].

Let G be a graph with vertex set V. For any subset S of V we define $N_G(S)$ to be the set of vertices v in $V \setminus S$ such that v is adjacent to some vertex of S.

Theorem 2.3.3 *Let* G *be a bipartite graph with partite sets* V_1 *and* V_2 *and let* d *be a positive integer. Then*

$$\beta_1(G) \geq |V_1| - d \tag{2.49}$$

if and only if for all subsets S *of* V_1,

$$|N_G(S)| \geq |S| - d. \tag{2.50}$$

Proof. Construct a bipartite graph G_d from G by adding d new vertices to V_2 and all possible edges between V_1 and the d new vertices. Then Hall's theorem is applied to G_d. □

As usual, our sample space in this section is $\Gamma(n, n)$ with the uniform probability distribution. Each tree T in $\Gamma(n, n)$ is a bipartite graph in which V_1 is the set of n light vertices and V_2 is the set of n dark vertices. A subset S of V_1 is *bad* if

$$|N_T(S)| < |S| - d. \tag{2.51}$$

By Theorem 2.3.3, if there are no bad sets in T, then

$$\beta_1(T) \geq n - d. \tag{2.52}$$

Let X_d be the random variable which counts the number of bad sets. We want to determine d as a function of n so that

$$E(X_d) \to 0 \text{ as } n \to \infty. \tag{2.53}$$

Then we can say "almost surely, there are no bad sets". Hence for almost all (n,n)-trees T in $\Gamma(n,n)$

$$\beta_1(T) \geq n - d. \tag{2.54}$$

Naturally we will try to make d as small as possible. We set

$$d = \alpha n, \tag{2.55}$$

where α is a constant between 0 and 1 .

Suppose S is a bad set of order k in a tree T. Since T is connected, we must have

$$d + 1 < k < n. \tag{2.56}$$

Let A_k be the set of all trees in $\Gamma(n,n)$ for which a specified k-subset of V_1 has all of its neighbors in a specified $(k-d-1)$-subset of V_2. Thus each tree in A_k has at least one bad set of order k. Then $|A_k|$ can be found by applying Corollary 2.1.1 with $n_1 = n - (k-d-1)$, $n_2 = n - k$, $n_3 = k - d - 1$ and $n_4 = k$. We find that

$$|A_k| = n^{n-k-1} n^{k-d-2} (n-k)^{n-(k-d-1)} (k-d-1)^k. \tag{2.57}$$

Then we have the following expression for the expected number of bad sets:

$$E(X_d) = \sum_{d+1<k<n} \binom{n}{k} \binom{n}{k-d-1} P(A_k), \tag{2.58}$$

where $P(A_k)$ is the ratio of (2.57) and n^{2n-2}.

We can now estimate $E(X_d)$ just as we handled the expectation in the previous section. First we apply Stirling's formula to the binomial coefficients. Then we simplify the summands and apply the ratio test. We find that the terms increase for

$$k < (n+d)/2,$$

and they decrease for

$$k \geq (n+d)/2.$$

Hence $E(X_d)$ is bounded by the product of the largest term, which occurs at $k = (n+d)/2$, and the length of the sum. Futher computation yields

$$E(X_d) = O(1)(2^{1-\alpha}(1-\alpha)^{2\alpha}/(1+\alpha)^{1+\alpha})^n. \tag{2.59}$$

This bound is seen, at once, to be identical to that of formula (2.35) in the previous section. On further examination of the formula (2.12) for $E(X_k)$ and (2.58) for $E(X_d)$, we see that these are virtually identical. One can be obtained from the other by an appropriate change in the names of the variables. Hence no improvement in Theorem 2.3.2 is possible from this approach.

2.4 *Comparison with random bipartite graphs*

In the previous two sections, we discussed bounds of β_0 and β_1 for almost all (n,n)-trees in $\Gamma(n,n)$. Following Bollobás (see [Bo85] p.52) we let $\mathcal{G}\{K(n,n);\, p\}$ denote the probability space of all bipartite labeled graphs with partite sets V_1 and V_2,

$$|V_1| = |V_2| = n,$$

in which each $V_1 - V_2$ edge is selected with probability p. More specifically for each graph G with M edges, the probability assigned to G is

$$P(G) = p^M(1-p)^{n^2-M}. \tag{2.60}$$

Now we seek bounds of β_0 and β_1 for almost all bipartite graphs in $\mathcal{G}\{K(n,n);\, p\}$. We preserve most of the approach and notation from the previous sections.

Fix $k > n$ and let X_k be the random variable that counts independent sets of order k. Then the expectation of X_k is

$$E(X_k) = \sum \binom{n}{i} \binom{n}{j} (1-p)^{ij}, \tag{2.61}$$

where the sum is over all i, j with

$$0 < i, j < n \text{ and } i + j = k. \tag{2.62}$$

Since the expected number of edges in a random graph in $\mathcal{G}\{K(n,n);\ p\}$ is pn^2, we choose

$$p = 2/n, \tag{2.63}$$

so that the number of edges will be approximately $2n - 1$, as in an (n,n)-tree.

We set

$$k = (1+\alpha)n, \tag{2.64}$$

as before where α is a positive constant, and we will try to determine α as small as possible so that

$$E(X_k) \to 0 \text{ as } n \to \infty.$$

Then we can say

$$\beta_0(T) \leq k = (1+\alpha)n, \tag{2.65}$$

for almost all bipartite graphs G in $\mathcal{G}\{K(n,n);\ p\}$.

As before in section 2.2, symmetry allows us to eliminate the variable j in formula (2.61) for $E(X_k)$. We find

$$E(X_k) \leq 2 \sum \binom{n}{i} \binom{n}{k-i} (1-p)^{i(k-i)}, \tag{2.66}$$

where the sum is over all i with

$$\alpha n < i < (n + \alpha n + 1)/2 \text{ and } p = 2/n. \tag{2.67}$$

We estimate $E(X_k)$ in (2.66) as usual. First we apply Stirling's formula to the binomial coefficients. Then simplify the summands to find

$$E(X_k) = O(1)n^{2n+1} \sum_{\alpha n<i<(n+\alpha n+1)/2} a_i, \tag{2.68}$$

where

$$a_i = (\frac{n-2}{n})^{i(1+\alpha)n-i^2} \frac{1}{i^{i+1/2}(n-i)^{n-i+1/2}(i-\alpha n)^{i-\alpha n+1/2}(n+\alpha n-i)^{n+\alpha n-i+1/2}}. \tag{2.69}$$

We make a slight change in notation by setting

$$i = \delta n \text{ for} \tag{2.70}$$

$$\alpha < \delta \leq (1+\alpha)/2.$$

Then

$$(\frac{n-2}{n})^{i(1+\alpha)n-i^2} = (\frac{n-2}{n})^{(\delta(1+\alpha)-\delta^2)n^2}. \tag{2.71}$$

Next we use the fact that

$$(\frac{n-2}{n})^{(\delta(1+\alpha)-\delta^2)n^2} < e^{-2(\delta(1+\alpha)-\delta^2)n}. \tag{2.72}$$

Then the equation (2.68) will become

$$E(X_k) = O(1)n^{-1} \sum_{\alpha n<\delta n<(n+\alpha n+1)/2} D(\alpha,\delta)^n, \tag{2.73}$$

where

$$D(\alpha,\delta) = 1/e^{2(\delta(1+\alpha)-\delta^2)}\delta^\delta(1-\delta)^{1-\delta}(\delta-\alpha)^{\delta-\alpha}(1+\alpha-\delta)^{1+\alpha-\delta}. \tag{2.74}$$

To make the right side of (2.73) approach zero, we need to find an α such that

$$D(\alpha,\delta) < 1, \tag{2.75}$$

for all

$$\alpha < \delta \leq (1+\alpha)/2.$$

A straightforward calculation shows that

$$D(.5, \delta) < 1 \text{ for all } .5 < \delta \leq .75. \tag{2.76}$$

Hence

$$E(X_k) = O(1)D(\alpha, \delta)^n \to 0 \text{ as } n \to \infty, \tag{2.77}$$

if

$$\alpha = .5. \tag{2.78}$$

On the other hand,

$$E(X_k) \geq c_1 n^{2n+1} a_t, \tag{2.79}$$

where c_1 is a positive constant and

$$t = (1+\alpha)n/2. \tag{2.80}$$

$$a_{(1+\alpha)n/2} \geq c_2 n^{-2n-2} A^n, \tag{2.81}$$

where c_2 is a positive constant and

$$A = \frac{4}{e^{(1+\alpha)^2/2}(1+\alpha)^{1+\alpha}(1-\alpha)^{1-\alpha}}. \tag{2.82}$$

And we find

$$D(.49, \delta) > 1 \text{ for some } \delta \text{ near}.7. \tag{2.83}$$

For example,

$$D(.49, .7) = 1.0258. \tag{2.84}$$

Which implies that the estimate

$$\alpha = .5$$

cannot be improved significantly by this method.

These observations are summerized in the following theorem.

Theorem 2.4.1 *For almost all bipartite graphs G, the vertex independence number $\beta_0(G)$ satisfies the inequality:*

$$n \leq \beta_0(G) \leq (1.5)n. \tag{2.85}$$

As expected, Theorems 2.2.1 and 2.4.1 give similar bounds on the independence number. But the upper bound in 2.4.1 is slightly higher than that of 2.2.1. This might be accounted for by the fact that the random graph model is less restricted. Hence there are more opportunities for large independent sets.

Bibliography

[AlSE92] N.Alon, J.H. Spencer and P. Erdös, *The Probabilistic Method,* Wiley-Interscience, New York (1992).

[Au60] T.L. Austin, The enumeration of point labelled chromatic graphs and trees, *Canad. J. Math.* **12** (1960) 535-545.

[Be74] E.A. Bender, Asymptotic methods in enumeration, *SIAM Rev.* **16** (1974) 485-515.

[Bo85] B. Bollobás, *Random Graphs,* Academic Press, London (1985).

[C89] A. Cayley, A theorem on trees, *Quart. J. Math.* **23** (1889) 376-378.

[Er47] P. Erdös, Some remarks on the theory of graphs, *Bull. Amer. Math. Soc.* **53** (1947) 292-294.

[Ga59] T. Gallai, Über extreme Punkt- und Kantenmengen, *Ann. Univ. Sci. Budapest, Eötvös Sect. Math.* **2** (1959) 133-138.

[Ha35] P. Hall, On representatives of subsets, *J. London Math. Soc.* **10** (1935) 26-30.

[H69] F. Harary, *Graph Theory,* Addison-Wesley, Reading (1969).

[Kn68] D. E. Knuth, Another enumeration of trees, *Canad. J. Math.* **20** (1968) 1077-1086.

[Kö31] D. König, Graphen und Matrizen, *Math. Fiz. Lapok.* **38** (1931) 116-119.

[M70] J. W. Moon, *Counting Labelled Trees,* Canad. Math. Congress, Montreal (1970).

[MeM73] A. Meir and J. W. Moon, The expected node-independence number of random trees, *Nederl. Akad. Wetensch. Proc. Ser. A* **76** = Indag. Math.**35** (1973) 335-341.

[MeM75] A. Meir and J. W. Moon, The expected node-independence number of various types of trees, *Recent Advances in Graph Theory,* Academia Praha (1975) 351-363.

[Pa85] E. M. Palmer, *Graphical Evolution,* Wiley-Interscience, New York (1985).

[Pa92] E. M. Palmer, Matchings in random superpositions of bipartite trees, *J. Comput. Appl. Math.* **41** (1992).

[Sch92] E. Schmutz, Matchings in Superpositions of (n,n)-Bipartite Trees, preprint.

[Sc62] H.I. Scoins, The number of trees with nodes of alternate parity, *Math. Proc. Cambridge Philos. Soc.* **58** (1962) 12-16.