# A Note on Matrices of Trees *

A. Ehrenfeucht
G. Rozenberg

CU-CS-044-74

University of Colorado at Boulder
DEPARTMENT OF COMPUTER SCIENCE

* This research was supported by the National Science Foundation under grant no. ACR-93-57812 and by the Department of Energy under grant no. DE-FG03-97ER25325.

A Note on Matrices of Trees*

by

A. Ehrenfeucht**

and

G. Rozenberg***

Report #CU-CS-044-74 July 1974

* This work supported by NSF Grant #GJ-660

** Department of Computer Science
University of Colorado
Boulder, Colorado 80302 U.S.A.

*** Institute of Mathematics
Utrecht University
Utrecht-De Uithof HOLLAND
and
Department of Mathematics
University of Antwerp, U.I.A.
Wilrijk BELGIUM

All correspondence to: G. Rozenberg
Institute of Mathematics
Utrecht-De Uithof
Holland

ABSTRACT

The notions of a matrix of trees and of a well formed matrix of trees are introduced. They arose from research in formal language theory. It is proved that each matrix of trees contains a "relatively large" submatrix which is well formed.

INTRODUCTION

One of the central notions of formal language theory is that of a derivation in a grammar (see e.g., Salomaa [2]). From a graph-theoretic point of view each derivation may be viewed as an arrangement of trees into a matrix. Using this approach we were able to discover a quite useful structure of derivations in the so-called deterministic ETOL systems (see Ehrenfeucht and Rozenberg [1]).

The notion of a <u>matrix of trees</u> which arose in this way is investigated in this paper. The main result presented here is used in a very essential way to prove the main result of Ehrenfeucht and Rozenberg [1].

DEFINITIONS AND EXAMPLES

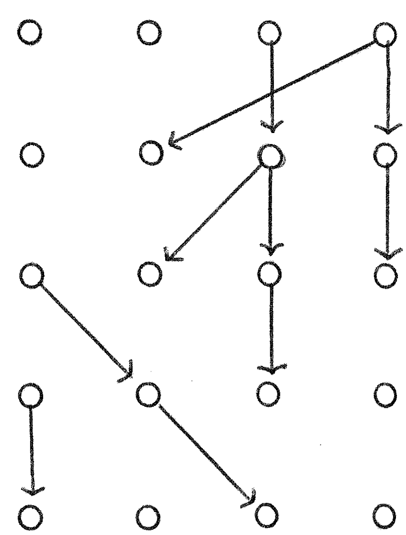
In this section we provide definitions (illustrated by examples) of the main notions used in this note.

Definition 1. A (n×k) matrix of trees (abbreviated as a (n×k) t-matrix) is a directed graph whose nodes form a (n×k) matrix which satisfies two conditions:

(i) each node in the graph has at most one ancestor,

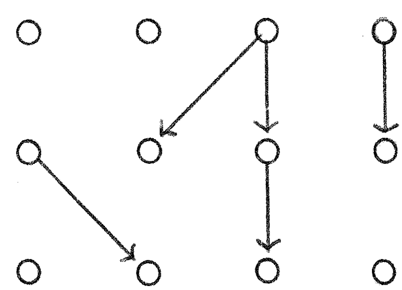
(ii) if there is an edge leading from node (i,j) to node $(\bar{i},\bar{j})$, for some $1 \leq i, \bar{i} \leq n$ and $1 \leq j, \bar{j} \leq k$, then $\bar{i} = i + 1$.

Example 1. The following is an example of a (5×4) t-matrix:



Definition 2. Let $G_1$ be a (n×k) t-matrix and let $G_2$ be a (m×k) t-matrix for some $m \leq n$. We say that $G_2$ is a sub-t-matrix of $G_1$ if the (m×k) matrix of nodes of $G_2$ is obtained by omitting some (maybe none) rows from the matrix of nodes of $G_1$ and there is an edge between two nodes in $G_2$ if and only if this edge is in the transitive closure of $G_1$.

Example 2. The (3×4) t-matrix shown below



is a sub-t-matrix of the t-matrix in Example 1. It is obtained by omitting the second and the fifth rows of the original matrix.

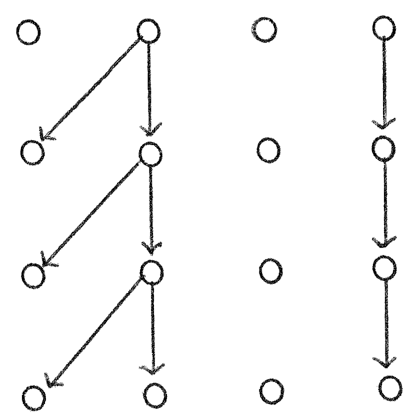
It should be clear to the reader that if $G_1$ is a sub-t-matrix of a t-matrix G and if $G_2$ is a sub-t-matrix of $G_1$, then $G_2$ is a sub-t-matrix of G.

Definition 3. A (n×k) t-matrix G is said to be well formed if it satisfies the following two conditions:

1) If a node (i,j) has descendants then (i+1,j) is one of them.

2) If there is an edge leading from (i,j) to (i+1,$\ell$), then, for every p in {1, ..., n-1}, there is the edge leading from (p,j) to (p+1,$\ell$).

Example 3. The t-matrices in examples 1 and 2 are not well formed. The following is an example of a well formed (5×4) matrix:



Given a t-matrix G we can in an obvious sense talk about its rows and columns.

Definition 4. Let G be a t-matrix and let $\alpha$ be one of its columns.

1) We say that $\alpha$ is a column of type 1 if from each node in $\alpha$ (except for the last one) there is an edge leading to the next node in $\alpha$.

2) We say that $\alpha$ is a column of type 2 if no node in $\alpha$ has a descendant and if for every node in $\alpha$ which has an ancestor, the ancestor belongs to a column of type 1.

3) We say that $\alpha$ is an arranged column if it is either of type 1 or of type 2.

Example 4. For the t-matrix of Example 3, the second and the fourth columns are of type 1 and the first and the third columns are of type 2.

It should be clear to the reader that if a column is arranged in a t-matrix G, then it "stays arranged" in all sub-t-matrices of G.

MAIN RESULT

In this section we shall prove that each t-matrix has a "relatively large" well formed sub-t-matrix.

Lemma 1. Let G be a $(n \times k)$ t-matrix which has $\ell$ arranged columns for some $1 \leq \ell \leq k-1$. There exists a sub-t-matrix $G_1$ of G such that

1) $G_1$ has at least $(\ell+1)$ arranged columns, and

2) $G_1$ is of order $(n_1 \times k)$ for some $n_1 \geq \frac{\sqrt{n}}{k}$.

Proof.

Let G satisfy the statement of the lemma.

Let $P_G$ be the collection of all paths in G the nodes of which do not belong to any of the arranged columns in G. Let $\gamma$ be a path from $P_G$ such that no path in $P_G$ is longer than $\gamma$.

(i) Let us assume that the length of $\gamma$ is larger than $\sqrt{n}$.

Let $C_\gamma$ be the set of columns which have at least one node that belongs to $\gamma$. Let $c_\gamma$ be a column from $C_\gamma$ such that no other column in $C_\gamma$ has more nodes in $\gamma$ than $c_\gamma$ has. Clearly $c_\gamma$ has at least $\frac{\sqrt{n}}{k}$ nodes in $\gamma$. Thus if we choose $G_1$ to be the sub-t-matrix of G consisting of all the rows of G having a node of $\gamma$ in their $c_\gamma$ column, then we see that $G_1$ satisfies conditions 1) and 2) and Lemma 1 holds.

(ii) Let us assume that the length of $\gamma$ is no larger than $\sqrt{n}$.

Now if we choose all the rows numbered† $1, \lceil\sqrt{n}\rceil, \lceil\sqrt{n}\rceil \cdot 2, \ldots$ we obtain a sub-t-matrix $G_1$ satisfying the statement of the Lemma.

This ends the proof of Lemma 1.

---

†For a real number r, $\lceil r \rceil$ denotes the smallest integer larger than r.

Lemma 2. Let G be a $(n\times k)$ t-matrix. Then there exists a sub-t-matrix H of G such that

1) all columns of H are arranged, and

2) H is of order $(m\times k)$ for some $m \geq \frac{n^{\frac{1}{2k}}}{k^k}$.

Proof

The result follows immediately from repeated (at most k times) application of lemma 1.

Now we can easily prove our main result, which, informally speaking, says that each t-matrix has a "relatively large" well formed sub-t-matrix.

Theorem. For every positive integer k there exist positive reals $r_k$ and $s_k$ such that for every positive integer n and for every $(n\times k)$ t-matrix G there exists a well formed $(m\times k)$ t-matrix H which is a sub-t-matrix of G and for which $m \geq r_k n^{s_k}$.

Proof.

Let k be a positive integer and let G be a $(n\times k)$ t-matrix. By Lemma 2, there exists a sub-t-matrix $\overline{H}$ of G of order $(p\times k)$, for some $p \geq \frac{n^{\frac{1}{2k}}}{k^k}$, in which each column is arranged.

Let $i,j\varepsilon\{1, \ldots, p-1\}$. We say that the $i^{th}$ and the $j^{th}$ rows of $\overline{H}$ are isomorphic if the following holds: for every $t_1$, $t_2$ in $\{1, \ldots, k\}$, there is an edge leading from $(i,t_1)$ to $(i+1,t_2)$ if and only if there is an edge leading from $(j,t_1)$ to $(j+1,t_2)$. By a type in $\overline{H}$ we mean a set of all of which are isomorphic rows of $\overline{H}$. Let T be a type in $\overline{H}$ such that no type in $\overline{H}$ contains more elements than T. As the number of different types is certainly no larger than $2^{k^2}$, the number of elements in T is at least $\frac{n^{\frac{1}{2k}}}{k^k} \cdot \frac{1}{2^{k^2}}$. If we choose now H to be the sub-t-matrix of $\overline{H}$ consisting of all these rows of $\overline{H}$ which are in T then it is clear that H is well formed. Consequently if we choose $r_k = \frac{1}{k^k 2^{k^2}}$ and $s_k = \frac{1}{2k}$ then H satisfies the statement of the theorem.

Thus the theorem holds.

ACKNOWLEDGEMENT

The authors would like to thank Professor Lloyd Fosdick for arranging a very enjoyable stay for the second author at the Computer Science Department at the University of Colorado at Boulder. Thanks to this arrangement the work on this and other papers could be continued.

REFERENCES

1. A. Ehrenfeucht and G. Rozenberg, On Structure of Derivations in EDTOL Systems, University of Colorado, Department of Computer Science, Technical Report No. CU-CS-046-74.

2. A. Salomaa, Formal Languagues, Academic Press, 1973.