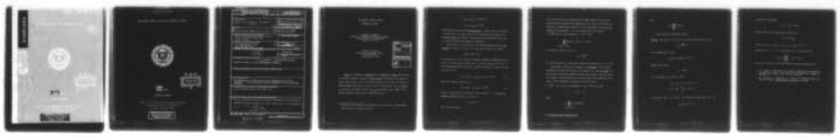
AD-A051 994 GEORGIA INST OF TECH ATLANTA SCHOOL OF INFORMATION A--ETC F/G 12/1
THE AVERAGE LENGTH OF PATHS EMBEDDED IN TREES.(U)
JAN 77 R A DEMILLO, R J LIPTON DAAG29-76-G-0338

UNCLASSIFIED GIT-ICS-77/03 ARO-14690.4-M NL

| OF |
AD
A051 994



END
DATE
FILMED
4-78
DDC

JAN.

GITICS-77-03

(Technical Report)

THE AVERAGE LENGTH OF PATHS EMBEDDED IN TREES

JAN ~~MARCH~~ 1977

TECHNICAL REPORT

School of Information and Computer Science
Georgia Institute of Technology
Atlanta, Georgia 30332

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) The Average Length of Paths Embedded in Trees,		5. TYPE OF REPORT & PERIOD COVERED (9) Interim Technical rept.,
		6. PERFORMING ORG. REPORT NUMBER (14) GIT-ICS-77/03
7. AUTHOR(s) (10) Richard A. DeMillo [redacted] Richard J. Lipton		8. CONTRACT OR GRANT NUMBER(s) (15) DAAG29-76-G-0338
9. PERFORMING ORGANIZATION NAME AND ADDRESS School of Information and Computer Science, Georgia Institute of Technology, Atlanta, GA 30332		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office, Post Office Box 12211, Research Triangle Park, NC 27709		12. REPORT DATE (11) Jan[redacted] 1977
		13. NUMBER OF PAGES five (5)
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (12) 8p.		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report) (18) ARO (19) 14690.4-M

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

array, embedding, graph, path length, tree.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Let A_n be defined so that the $n \times n$ array is embeddable in binary trees by dilating average path length by at most a factor of A_n. It is shown that as n approaches infinity the limit of $A_n = 0$.

$$\lim_{n \to \infty} A_n = 0.$$

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

410 044

THE AVERAGE LENGTH OF PATHS EMBEDDED IN TREES*

Richard A. DeMillo
School of Information and Computer Science
Georgia Institute of Technology
Atlanta, Ga. 30332

Richard J. Lipton
Department of Computer Science
Yale University
New Haven, Ct. 06520

A graph, G, consists of vertices V(G) and edges E(G); paths are sequences of vertices connected by edges, and path length is defined by the number of edges along the path. For $x,y \in V(G)$ we use $d_G(x,y)$ to denote the length of a minimal length path between x and y, if such a path exists. An $n \times n$ array, G_n, consists of vertices $V(G_n) = \{x_{ij}\}_{i,j \leq n}$ and edges which, except at the obvious extremal conditions, are linked as follows:

* The work of both authors was supported in part by the U.S. Army Research Office, Grant No. DAA G29-76-G-0338.

$$(x_{i,j}, x_{i+1,j}) \in E(G_n), \text{ and}$$

$$(x_{i,j}, x_{i,j+1}) \in E(G_n).$$

Such graphs are also called rook-connected. A binary tree is as defined in [1,2]; that is, a binary tree H is a connected acyclic graph with a designated root and ancestor - descendent relation defined so that each $x \in V(H)$ has at most two immediate descendents.

Let us write $G \leq_T H$ when there is a one-one mapping (called an embedding of G into H $\Phi : V(G) \to V(H)$, such that for all $(x,y) \in E(G)$,

$$d_H(\Phi(x), \Phi(y)) \leq T.$$

As described in [1], it follows from simple volumetric arguments that for all $T > 0$, there exists a binary tree H such that $H \not\leq_T G_n$, for all $n \geq 1$. The corresponding intuition for $G_n \leq_T H$ does not hold. It would now seem that since in G_n

$$|\{x \in V(G_n): \; d_{G_n}(x,y) \leq k\}| = O(k^2) \qquad (1)$$

while in a complete binary tree H

$$|\{x \in V(H): \; d_H(x,y)\}| \geq 2^{k-1} \qquad (2)$$

that $G_n \leq_T H$ would now be possible for some bounded T. It is therefore somewhat surprising that $G_n \leq_T H$ only if

$$T \geq \log n - 1.5$$

(See [1], for details).

It is still obvious from inspection that neighborhoods in trees can be much more densely growing than neighborhoods in arrays, and therefore by choosing a suitably global measure of loss of proximity, this difference should be distinguishable. In [2] we considered such a measure:

$G \leq^{edge}_{A} G^*$ if for some embedding $\Phi : V(G) \to V(G^*)$

$$\sum_{(x,y) \in E(G)} d_{G^*}(\Phi(x), \Phi(y)) \leq A \mid E(G) \mid .$$

It follows [2] that for b = 8.5

$$G_n \leq^{edge}_{b} H$$

for some binary tree H. This upper bound can be improved to $b = 7 - o(1)^{\dagger}$

The relation $\leq^{edge}_{A}$ may be thought of as averaging - with relative frequencies uniformly distributed to the edges E(G) - over the edges of G. We now make a more global definition which finally may be used to recover our original, although imprecise, intuitions about path lengths in binary trees. We will essentially average over shortest paths:

$G \leq^{paths}_{A} G^*$ if one is an embedding $\Phi : V(G) \to V(G^*)$ such that

$$\Gamma_n \leq A \cdot \Delta_n$$

where

$$\Gamma_n = \sum_{\Phi(x),\Phi(y)} d_{G^*} (\Phi(A,\Phi(y))$$

† L. Snyder, private communication.

and

$$\Delta_n = \sum_{x,y} d_G(x,y).$$

We then have the following theorem.

Theorem. For each $n \geq 0$, let A_n be the least real number such that

$$G_n \leq^{\text{paths}}_{A_n} H,$$

for a binary tree H. Then

$$\lim_{n \to \infty} A_n = \lim_{n \to \infty} \Gamma_n / \Delta_n = 0.$$

Proof we first show

$$\Delta_n = \Omega(n^5)$$

Let us choose $B_1, B_2 \subseteq V(G_n)$ so that

$$B_1 = \{x_{ij} : 1 \leq i, j \leq n / 4\}$$

$$B_2 = \{x_{ij} : \frac{3n}{4} \leq i,j \leq n\}$$

so that $|B_1 \times B_2| = [n^2 / 16]^2$. Now clearly, for any $(x,y) \in B_1 \times B_2$

$$d_{G_n}(x,y) \geq n / 2,$$

and hence by definition

$$\Delta_n \geq n^5 / 512 = \Omega(n^5)$$

We now obtain the following upper bound for Γ_n

$$\Gamma_n = O(n^4 \log n).$$

As in [2] let $A_{ij} \subseteq V(G_n)$, $1 \leq i, j \leq 2$, $|A_{ij}| = n^2 / 4$,

Denote the $n / 2 \times n / 2$ decomposition of G_n and notice that

$$\Gamma(n) \leq 4 \Gamma\left(\frac{n}{2}\right) + \frac{1}{2} n^4 \log n.$$

Thus $\Gamma(n) \leq \alpha n^4 \log n + \beta n^4$ from which the theorem follows directly.

1. R.J. Lipton, S. Eisenstat, R.A. DeMillo, "Space and Time Hierarchies for Classes of Control Structures and Data Structures", Journal of the ACM, Vol. 23, No. 4, Oct. 1976, pp. 720-732.

2. R.A. DeMillo, S.C. Eisenstat, R.J. Lipton, "Preserving Average Proximity in Arrays", Communications of the ACM (to appear).