MATHEMATICS OF OPERATIONS RESEARCH
Vol. 10, No. 4, November 1985
Printed in U.S.A.

ON SHORTEST PATHS IN GRAPHS WITH RANDOM WEIGHTS*†

REFAEL HASSIN‡ AND EITAN ZEMEL§

We consider the shortest paths between all pairs of nodes in a directed or undirected complete graph with edge lengths which are uniformly and independently distributed in [0, 1]. We show that the longest of these paths is bounded by $c \log n/n$ almost surely, where c is a constant and n is the number of nodes. Our bound is the best possible up to a constant. We apply this result to some well-known problems and obtain several algorithmic improvements over existing results. Our results hold with obvious modifications to random (as opposed to complete) graphs and to any distribution of weights whose density is positive and bounded from below at a neighborhood of zero. As a corollary of our proof we get a new result concerning the diameter of random graphs.

1. Introduction. There has been a growing interest in recent years in probabilistic analysis of optimization problems and algorithms. These include both "easy" problems, for which polynomial algorithms are available (e.g., Bloniarz 1983; Karp 1980; Rohlf 1978; Spira 1973; Walkup 1979; Weide 1980) and "hard" ones, for which such an algorithm is unlikely to exist (e.g., Cornuejols et al. 1980; Fischer and Hochbaum 1980; Hochbaum 1979; Karp 1977, 1979; Lueker 1981; Marchetti-Spaccamela et al. 1982; Papadimitriou 1981; Zemel 1982).

In this paper, we analyze the problem of finding the shortest paths between pairs of nodes in a directed or undirected complete graph whose edge lengths are uniformly and independently distributed in [0, 1]. Our main theorem states that there exists a constant c such that the distance between each pair of nodes is bounded by $c \log n/n$ almost surely. We also show that the order of magnitude of this bound cannot be improved. The result is then applied to a variety of situations to yield some improvements in the algorithmic performance. These include shortest path problems, minimum ratio problems, location problems, etc.

We devote the next section to the statement and proof of our main theorem. Subsequent sections consider some of the applications.

2. Main theorem. Consider a complete graph $G_n = (V, E)$ with node set $V = \{v_1, \ldots, v_n\}$ and edge lengths $d(v_i, v_j) \equiv d_{ij}$. We consider here the undirected case $d_{ij} = d_{ji}$. However, all our results apply without any modification for the directed case as well. Let D_{ij} denote the length of the shortest path between a given pair of nodes v_i and v_j. We are interested in the size of D_{mx}, the maximum, over all pairs, of D_{ij} when the individual edge lengths, d_{ij}, are uniformly and independently distributed in the interval [0, 1]. Our main result is the following:

THEOREM. *There exists a constant c such that* $D_{mx} \leqslant c < \log n/n$ *almost surely.*

*Received January 27, 1983; revised April 30, 1984.
AMS 1980 subject classification. Primary: 05C38.
IAOR 1973 subject classification. Main: Graphs.
OR/MS Index 1978 subject classification. Primary: 489 Networks/graphs:theory.
Key words. Random graphs, probabilistic algorithms, shortest path.
†Supported in part by NSF Grant ECS-812141 and by the Mia Fischer David Foundation through the Israel Institute of Business Research, Tel Aviv University.
‡Tel Aviv University
§Northwestern University and Tel Aviv University.

0364-765X/85/1004/0557$01.25

The technical meaning of the statement "almost surely" (a.s.) is the following. For each graph size, n, let p_n denote the probability that a certain assertion (such as the one expressed in the theorem) does not hold. We say that the assertion holds *almost surely* if $\sum_{n=1}^{\infty} p_n < \infty$. Obviously, this condition is stronger than the requirement that p_n tends to zero for large n. When only the latter condition holds, we say that the assertion holds in *probability*. On the difference between the two concepts and the relevance of the stronger one to optimization problems, see Rényi (1970) and Steele (1981).

We include in this section also a proof that the order of magnitude claimed by our theorem cannot be improved, i.e., we exhibit a constant c_1 such that $\Pr[D_{mx} \geq c_1 \log n/n]$ is considerable. As will be apparent to the reader, we do not derive the sharpest possible values for c and c_1, and thus there is a considerable gap between them. This gap, however, does not affect the applications we discuss here.

For every two functions $f(n)$, $g(n)$, we use the standard asymptotic notation $g(n) = o(f(n))$ if $\operatorname{Lim}_{n\to\infty} g(n)/f(n) = 0$, $g(n) = \Omega(f(n))$ if $f(n) = o(g(n))$ and $g(n) = O(f(n))$ if $\lim g(n)/f(n) = c$ for some constant c.

PROOF OF THE THEOREM. The general thrust of the proof is as follows. Let $b > 0$ be a constant to be specified later. For each node $v_i \in V$, we construct a set $H_i \subseteq V$ such that almost surely $|H_i| \geq (n \log n)^{1/2}$, and $D_{ik} \leq b \log n/n$ for each $v_k \in H_i$. Clearly, if $H_i \cap H_j \neq \emptyset$, then $D_{ij} \leq 2b \log n/n$. Otherwise, we show that there almost surely exists an edge (k, l) such that $v_k \in H_i$, $v_l \in H_j$ and

$$D_{ij} \leq D_{ik} + D_{lj} + d_{kl} \leq c \log n/n$$

for a constant c which is essentially equal to $2b$.

We first examine the construction of H_i. Our construction is sequential; that is, we generate a nested sequence of subsets $H_i^0 \subseteq H_i^1 \subseteq H_i^2 \subseteq \cdots \subseteq H_i^r \equiv H_i$ and demonstrate that $H_i = H_i^r$ has the required properties.

We open with a useful approximation to the Binomial distribution. It is based on Bernstein's inequality, and its proof can be found in Rényi (1970, p. 210).

LEMMA 1. *Let* $0 < p < 1$, $q = 1 - p$, $x < \frac{1}{2}\sqrt{n/pq}$. *Then*

$$\sum_{|r-np| > x\sqrt{npq}} \binom{n}{r} p^r q^{n-r} \leq 2\exp(-x^2/4).$$

Let α be a given constant, $q = (\alpha - 1)^2/4\alpha$. Define $H_i^0 = \{v_j \in V \mid d_{ij} \leq \alpha \log n/n\}$. Then it follows from Lemma 1 that $p_i \equiv \Pr[|H_i^0| < \log n] \leq 2n^{-q}$. For $q > 2$, which can be achieved by $\alpha > 10$, we get $p_i = o(1/n^2)$. Thus we have shown:

LEMMA 2. *For* $\alpha > 10$, *the* n *inequalities* $|H_i^0| \geq \log n$, $i = 1, \ldots, n$, *hold simultaneously almost surely.*

Pick a particular vertex $v_i \in V$ and let H_i^0 be the set of Lemma 2. We now show how H^{l+1} can be constructed from H_i^l for $l = 0, \ldots, r-1$ until we obtain the set $H_i \equiv H_i^r$ with the desired properties. Specifically, let $H_i^{l+1} = H_i^l \cup F_i^{l+1}$ where $F_i^{l+1} \cap H_i^l = \emptyset$. For convenience we define $F_i^0 \equiv H_i^0$.

Let β be a constant to be specified later, and let $r_l = \alpha \log n/n + (l-1)\beta/n$, $l = 1, \ldots, r$ with $r_0 \equiv 0$.

We assume that for $m = 0, \ldots, l$, F_i^m is such that for each $v_k \in F_i^m$, a particular path exists from v_i to v_k of length D_i^k which satisfies $r_m < D_i^k \leq r_{m+1}$. This assumption clearly holds for $m = 0$ since by construction, for each $v_k \in H_i^0$, $0 < d_{ik} \leq \alpha \log n/n \equiv r_1$ and thus we can take $D_k^i = d_{ik}$. To see how this property can be extended to

$m = l + 1$, let, for each $v_k \in H_i^l$,

$$\Delta_{kl}^i = \left(r_{l+1} - D_i^k, r_{l+2} - D_i^k\right].$$

Now let

$$F^{l+1} = \left\{ v_j \notin H_i^l \mid d_{kj} \in \Delta_{kl}^i \text{ for some } v_k \in H_i^l \right\}.$$

In words, f_i^{l+1} consists of those nodes v_j currently *not* in H_i^l but which can be reached from some node $v_k \in H_i^l$ by an edge whose length $d_{kj} \in \Delta_{kl}^i$. This makes for a path from v_i to v_j (via v_k) of total length $D_i^j = D_i^k + d_{kj}$ which satisfies the required condition: $r_{l+1} < D_i^j \leqslant r_{l+2}$. The reader may recall that, as per Lemma 2, the size of H_i^0 is at least $\log n$ almost surely. Similarly, Lemma 3 below indicates that F_i^{l+1} (and hence H_i^{l+1}) cannot be "too small" in relation to H_i^l:

LEMMA 3. *Let γ be a positive constant such that $\gamma < \beta$. Set*

$$t = \left[\beta - \gamma - h(\beta^2 + 2)/2n\right]^2/4\beta \quad \textit{and} \quad h = |H_i^l|.$$

If $h < (n \log n)^{1/2}$, then $\Pr[|F_i^{l+1}| < \gamma h] \leqslant 2n^{-t}$.

PROOF. First note that $h = |H_i^l| \geqslant |H_i^0| \geqslant \log n$. Next, for each pair $v_k \in H_i^l$, $v_j \notin H_i^l$ let P_{kj} denote the probability of the event "$d_{kj} \in \Delta_{kl}^i$" *conditioned* on $v_k \in H_i^l$, $v_j \notin H_i^l$. Clearly, these events are mutually independent for the various choices of j and k. Also, for each such pair, we have $p_{kj} \geqslant \beta/n$. This is due to the fact that Δ_{kl}^i has length β/n and that Δ_{kl}^i is disjoint from any of the intervals Δ_{km}^i, $m < l$, which may have been examined before. Thus, the number of elements of F_i^{l+1} is a binomial random variable with probability of "success" at least $1 - (1 - \beta/n)^h$ and the number of trials is $n - h = n(1 - h/n)$. The expected value of the number of elements in F_i^{l+1} is at least

$$(n-h)\left[1 - \left(1 - \frac{\beta}{n}\right)^h\right] \geqslant n\left(1 - \frac{h}{n}\right)\frac{\beta h}{n}\left(1 - \frac{1}{2}\,\frac{\beta h}{n}\right) \geqslant \beta h\left(1 - \left(\frac{\beta}{2} + 1\right)\frac{h}{n}\right)$$

and the required result now follows directly from Lemma 1.

Note that for large values of n, t is essentially equal to $(\beta - \gamma)^2/4\beta$.

Call an iteration, l, a "success" if $|H^{l+1}| \geqslant (1 + \gamma)|H^l|$ or if $H^{l+1} \geqslant (n \log n)^{1/2}$. Clearly, after at most $\log n/(2\log(1 + \gamma)) \equiv r^*$ successes, H_i^{l+1} achieves the desired size of $(n \log n)^{1/2}$. By Lemma 3, the probability of success at each iteration is at least $1 - 2n^{-t}$. Let δ be a positive constant to be specified later and let

$$r_0 = (1 + \delta)\log n/(2\log(1 + \gamma)), \qquad s = \left[\delta - \frac{2}{n^t}(1 + \delta)\right]^2/8(1 + \delta)\log(1 + \gamma).$$

LEMMA 4. *Let r be an integer, $r \geqslant r_0$, then* $p = \Pr[|H_i^r| < (n \log n)^{1/2}] \leqslant 2n^{-sn^t}$.

PROOF. p is bounded from above by the probability that r independent trials with probability of success $1 - 2n^{-t}$ each yield less than r^* successes. The desired result follows directly from Lemma 1 by substitution.

We have thus shown that the n inequalities $|H_i| \equiv |H_i^r| \geqslant (n \log n)^{1/2}$, $i = 1, \ldots, n$, hold simultaneously almost surely. By our construction method

$$D_{ik} \leqslant D_i^k \leqslant \alpha \log n/n + r\beta/n \equiv b \log n/n$$

for every $v_k \in H_i$. We now show that if $H_i \cap H_j = \emptyset$, we can find a sufficiently short edge connecting H_i and H_j which makes for a path from v_i to v_j of length $c \log n/n$.

For each pair $v_k \in H_i$, $v_m \in H_j$, let $\alpha_{km} = \min\{d_i^k, D_j^m\}$. Clearly $\alpha_{km} \leqslant b \log n/n$. Let

$$\hat{\Delta}_{km} = (b \log n/n - \alpha_{km}, b \log n/n - \alpha_{km} + \mu/n]$$

where $\mu > 0$ is a constant to be specified later. Note that the length of $\hat{\Delta}_{km}$ is μ/n. Also $\hat{\Delta}_{km}$ is disjoint from any of the intervals $\Delta_{kl}^i, \Delta_{ml}^j$, $l = 0, \ldots, r$, examined throughout the construction of H_i and H_j since the largest number contained in any of these intervals is $b \log n/n - \alpha_{km}$. Thus, the events $d_{km} \in \hat{\Delta}_{km}$, conditioned on $v_k \in H_i$, $v_m \in H_j$, $H_i \cap H_j = \emptyset$, are mutually independent and each has a probability of occurrence at least μ/n. There are $n \log n$ events of this type (one for each edge connecting H_i and H_j) so that the probability that none of these actually occurs is at most $n^{-\mu}$. For $\mu > 3$, this implies that for all pairs v_i, v_j, such that $H_i \cap H_j = \emptyset$ *simultaneously*, an appropriate edge can be found almost surely. Thus, the $n(n-1)/2$ inequalities $D_{ij} < (2b + \mu/\log n)\log n/n$ hold simultaneously almost surely. This proves the theorem since for large values of n, $2b + \mu/\log n$ is essentially equal to $2b$ and in any case can be bounded by a constant c.

The constant c of our theorem is rather large. For instance, by choosing $\alpha = 10$, $\beta = 2$, $\gamma = 1/2$, $\delta = 1$, we get that b is roughly equal to 13.33 so that c can be taken as 27. A sharper analysis can reduce this constant dramatically, perhaps by a factor of 10. Nevertheless, the order of magnitude cannot be improved:

LEMMA 6. $\Pr[D_{mx} \geqslant \log n/n] \geqslant e^{-1}$.

PROOF. A random graph with probability of each edge $p = \log n/n$ is not connected with probability e^{-1} (see Erdös and Spencer 1974, Chapter 16).

In the following sections, we proceed to examine some of the algorithmic implications of our theorem.

3. Shortest paths and spanning tree problems. Here and in the following sections we consider a complete (directed or undirected) graph $G = (E, V)$ with n nodes and a function d which assigns a length $0 \leqslant d_{ij} \leqslant 1$ to every edge $(v_i, v_j) \in E$. The values of d_{ij} are assumed to be independent uniform random variables. The main idea is that since all distances D_{ij} are a.s. bounded by $c \log n/n$, then in many problems edges which are larger than $k \log n/n$ for some constant k, will almost surely not be used in an optimal solution. Thus, in $O(n^2)$ preprocessing time, these edges can be deleted and be excluded from further consideration. This operation leaves us with a random graph $\hat{G} = (V, \hat{E})$ such that for every $(v_i, v_j) \in E$, $\Pr[(v_i, v_j) \in \hat{E}] = k \log n/n$, and $|\hat{E}| = O(n \log n)$ almost surely. Then a standard algorithm can be applied to $\hat{G}$ to find an optimal algorithm in reduced effort.

We note that by examining the solution obtained on $\hat{G}$, it should be possible to check whether the exclusion of the "long" edges was, in fact, justified. If not, the procedure failed, and the standard method must be applied to G in order to obtain the correct optimal solution. Thus, the worst case bound is identical with that of the standard method, but the algorithm's worst case (or average) performance time is almost surely very close to its worst case (or average) performance time on $\hat{G}$.

Since $O(n^2)$ is required just to scan the edges of G to obtain $\hat{G}$, there is no point in applying this method to problems which already have an $O(n^2)$ algorithm such as finding the shortest path between a given pair of nodes or the minimum length spanning tree. However, the method can be used to expedite algorithms whose running time is longer.

A case in point are shortest paths problems when more than one pair is involved. Note that our method requires $O(n^2)$ processing time after which, using Fredman and Tarjan (1984), all the shortest paths from a given node v_s to the other vertices of G can be computed in $O(n \log n)$ almost surely. An alternative procedure which can be used here is Bloniarz (1983), which is based on Spira's (1973) method and which can find all shortest paths leaving v_s in $O(n \log n \log^* n)^1$ expected time ($O(n^2)$ worst case time), after a preprocessing phase which requires $O(n^2 \log n)$. Thus, the order of running time of our algorithm almost surely is better than Bloniarz's order of *expected* time. Note, however, that the latter method is valid under less restrictive probabilistic assumptions than ours.

Suppose next that we are interested in finding a shortest $v_s - v_t$ path for some given $v_s, v_t \in V$, containing at most P edges. The problem can be solved in $O(mP)$ time for a graph with m edges using dynamic programming. Since the restriction on the number of edges makes the shortest allowed path larger, it is not always possible to restrict our search to $\hat{G}$. However, examination of the proof in §2 shows that for any pair $v_i, v_j \in V$ the number of edges required to construct the $v_i - v_j$ path is almost surely less than $(1 + \delta)\log n/\log(1 + \gamma) + 3$ which can be very generously bounded by $4 \log n$. Thus, the shortest $v_s - v_t$ path containing no more than P edges for $P \geqslant 4 \log n$ can be found almost surely in $O(Pn \log n + n^2)$ as opposed to $O(n^2 P)$ which is the regular bound.

As we have already mentioned, application of our theorem to the minimum spanning tree problem does not improve the complexity bound since this problem already has $O(n^2)$ algorithms. However, our method may be useful when a sequence of minimum spanning trees is to be computed. One example is a generalization of the minimum spanning tree problem, called the Steiner network problem. Another will be given in §5. Let the number of nodes of the graph be n, and suppose that a specified set of $n - s$ nodes is to be spanned by a tree of minimum weight. While this problem is known to be NP complete, Lawler (1976) presents an algorithm which for a fixed value of s is polynomial in n. The algorithm requires first the solution of all-pair shortest paths and then computes $O(2^s)$ solutions of minimum spanning trees on subgraphs with no more than $2(n - s - 1)$ nodes. Thus, its worst case complexity is $O(n - s)^2 2^s + n^3)$.

By our theorem, almost surely all distances between nodes in the graph are shorter than $c \log n/n$. In $O(n^2)$ preprocessing, all edges larger than this value can be deleted and the algorithm can be restricted to the resulting graph which almost surely has $O(n \log n)$ edges. Using any $O(m \log \log n)$ algorithm for the minimum spanning tree (e.g., Yao 1975, Cheriton and Tarjan 1976), we obtain a bound of

$$O\big(n \log n \log \log(n - s)2^s + n^2 \log n \log^* n\big)$$

almost surely.

4. Absolute P-center. For a set of points X on G and $v_i \in V$, let $D(i, X) = \min\{D_{ix} \mid x \in X\}$. The problem is to find the "weighted absolute p-center", $X = \{x_1, \ldots, x_p\}$ and the "p-radius" r_p for which $r_p = \min_{|X|=p}\{\max_{v_i \in V} w_i D(i, X)\}$, where w_i are given weights.

By our theorem, $r_p \leqslant c \log n/n$ almost surely. Thus, if $d_{ij} > 2c \log n/n$, then edge (v_i, v_j) almost surely contains no points in the optimal set X. This reduces the set of candidate edges to the $O(n \log n)$ edges which are shorter than $2c \log n/n$.

Kariv and Hakimi (1979) showed that the p-center problem on a general network is NP-hard. However, for $p = 1$, they give an algorithm which requires $O(n^3 \log n)$ time on complete graphs (and $O(n^3)$ time if $w_i = 1$ for all $v_i \in V$). The algorithm computes in $O(n \log n)$ time the best point on a given edge and this step is repeated n^2 times. This dominates the $O(n^3)$ time required by the algorithm to compute all-pair shortest distances in the graph.

[1] $\log^* n = \min\{i \mid \log^i n \leqslant 1\}$ and $\log^i$ denotes the ith iterate of the logarithm function.

It follows from our theorem that only $O(n \log n)$ best points need to be found, and thus the overall time of this step is $O(n^2 \log^2 n)$. The same amount of time is required to compute the all-pair shortest distances. Thus, for graphs with random edge lengths, the weighted one-center problem can be solved in just $O(n^2 \log^2 n)$ time, almost surely.

5. Minimum ratio problems. For any given feasible set $D \subseteq R^n$ consider:
Problem A: minimize $Z = \sum_E c_{ij} x_{ij}$ s.t. $x = (x_{ij} : (v_i, v_j) \in E) \in D$.
Problem B: minimize $R = (\sum_E a_{ij} x_{ij}) / (\sum_E b_{ij} x_{ij})$ s.t. $x \in D$.
Suppose that a function $g(n)$ is known such that when c_{ij} are uniform independent r.v.'s on the unit interval, then an optimal solution to A almost surely does not use edges with $c_{ij} > g(n)$. Suppose also that a_{ij} and b_{ij}, $(v_i, v_j) \in E$, are all independent uniform random variables on the unit interval. Let R^* denote the optimal value of R.

LEMMA. *For every $\epsilon > 0$, $0 \leqslant R^* \leqslant (1+\epsilon)^2 g(n)$ almost surely. In particular $R^* \leqslant 4g(n)$ almost surely.*

PROOF. We generate a solution $x \in D$ as follows: Let $\hat{G}$ be the graph obtained from G after deleting all edges with either $a > (1+\epsilon)g(n)$ or $b < (1+\epsilon)^{-1}$. Then the probability that a given $(v_i, v_j) \in E$ is in $\hat{G}$ is $g(n)$. By our assumption, this probability a.s. guarantees the existence of a feasible solution. Such a solution has however

$$R \leqslant \frac{(1+\epsilon)g(n)}{(1+\epsilon)^{-1}} = (1+\epsilon)^2 g(n).$$

It is a standard trick to solve ratio problems such as Problem B by solving a parametric series of problems of type A with costs $c_{ij} = a_{ij} - t b_{ij}$. The search terminates when $Z = 0$, in which case $t = R^*$. Obviously, if a bound on R^* is known, the values of t can be restricted to obey this bound. Clearly, when costs $c_{ij} = a_{ij} - t b_{ij}$ are considered a better solution is found than if $c_{ij} = a_{ij}$. Thus, edges with $a_{ij} - t b_{ij} > g(n)$ can be deleted. In particular, edges with $a_{ij} > 5g(n)$ can be deleted since $a_{ij} - t b_{ij} \geqslant a_{ij} - t \geqslant a_{ij} - 4g(n) \geqslant g(n)$. Thus, a.s. only $O(n \log n)$ edges must be considered. For example, consider the minimum ratio spanning tree problem which can be solved in $O(m \log^2 n \log\log n)$ time for a graph with m edges (Megiddo 1981) and in $O(m \log(nd) \log\log n)$ where d is the maximal edge length, expressed as integer (Zemel 1981). It follows from our main Theorem that $g(n) = cn \log n$ also for the MST Problem. If we replace m by $O(n \log n)$ we obtain $O(n \log^3 n \log\log n)$ which is dominated by the $O(n^2)$ preprocessing time. Thus the resulting algorithm requires $O(n^2)$ time almost surely.

6. Concluding remarks. Throughout the paper, we assumed that the edge weights are uniformly distributed. However, examination of our proof in §2 shows that the only important property of this distribution is that its density is positive and bounded from below at a neighborhood of zero. Thus, it is possible to extend our main theorem to other probability distributions as well.

Consider next a random graph $G_r = (E_r, V)$ such that for each $e \in E$, $\Pr[e \in E_r] = p_n$, and these events are independent. Then a proof parallel to that of §2 shows that $D_{mx} \leqslant c \log n / np_n$ almost surely. The applications of §§3–5 apply also in this case where edges larger than $c \log n / np_n$ can be excluded. Since $|E_r| = O(n^2 p_n)$ it follows that the resulting graphs have almost surely $O(n \log n)$ edges. The complexity of the proposed algorithm is thus unchanged except that the preprocessing effort is bounded by $O(m) = O(n^2 p_n)$, almost surely.

For example, the problem of finding a minimum spanning tree can be solved in $O(m \log\log n)$ time (see Yao 1975). After removing the "long" edges from G_r, we have $m = O(n \log n)$ edges and the problem can be solved on the resulting graph in

$O(n \log n \log\log n)$ time. If $P_n = \Omega(\log n \log\log n/n)$ then $m = \Omega(n \log n \log\log n)$ almost surely, and the preprocessing time is dominating, so that we obtain an $O(m)$ algorithm.

Another example concerns finding a shortest $v_s - v_t$ path. This can be done in $O(m \log n)$ time. After removing long edges we can thus solve the problem in $O(n \log^2 n)$. If $p_n = \Omega(\log^2 n/n)$, then $m = \Omega(n \log^2 n)$ a.s., and the preprocessing, which requires $O(m)$ time, dominates the overall complexity of the algorithm. We note that for graphs with $m = \Omega(n^{1+\epsilon})$ there already exist $O(m)$ algorithms for both the minimum spanning tree and shortest path (with nonnegative weights) problems (see Johnson 1975, 1977, Cheriton and Tarjan 1976).

A final result concerns the diameter of random graphs. Its proof follows easily from the proof of our theorem. Let b and r be as in §2.

COROLLARY. *Let $\hat{G} = (\hat{E}, V)$ be a random graph with $p_n \equiv \Pr[(v_i, v_j) \in \hat{E}] > b \log n/n$, then the diameter of $\hat{G}$ is almost surely less than or equal to $2r + 3$.*

Note that $2r < 4 \log n$ so that $p_n > b \log n/n$ suffices to guarantee $\dim(\hat{G}) = O(\log n)$. Very strong theorems concerning the diameter of random graphs appear in Bollobás (1980). However, they apply to different domains of p_n and $\dim(\hat{G})$.

References

Bloniarz, P. A. (1983). A Shortest Path Algorithm with Expected Time $O(n^2 \log n \log^* n)$. *SIAM J. Comput.* **13** 588–600.

Bollobás, B. (1981). The Diameter of Random Graphs. *Trans. Amer. Math. Soc.* **267** 41–52.

Cheriton, D. and Tarjan, R. E. (1976). Finding Minimum Spanning Trees. *SIAM J. Comput.* **5** 724–742.

Cornuejols, G., Nemhauser, G. L. and Wolsey, L. A. (1980). Worst Case and Probabilistic Analysis of Algorithms for a Location Problem. *Oper. Res.* **28** 847–858.

Erdös, P. and Spencer, J. (1974). *Probabilistic Methods in Combinatorics*. Academic Press, New York.

Fischer, M. L. and Hochbaum, D. S. (1980), Probabilistic Analysis of the Euclidean K-median Problem. *Math. Oper. Res.* **5** 27–34.

Fredman, M. L. and Tarjan, R. E. (1984). Fibonacci Heaps and Their Uses in Improved Network Optimization Algorithms. *25th Annual Sympos. on FOCS*. Singer Island, Florida, 338–346.

Hochbaum, D. S. (1979). The Probabilistic Asymptotic Properties of Some Combinatorial Geometric Problems. Manuscript, Carnegie-Mellon University.

Johnson, D. B. (1975). Priority Queues with Update and Finding Minimum Spanning Trees. *Inform. Process. Lett.* **4** 53–57.

———. (1977). Efficient Algorithms for Shortest Paths in Sparse Networks. *J. Assoc. Comput. Mach.* **24**, 1–13.

Kariv, P. and Hakimi, S. L. (1979). An Algorithmic Approach to Network Location Problems. Part 1: The p-Centers. *SIAM J. Appl. Math.* **37** 513 – 538.

Karp, R. M. (1977). The Probabilistic Analysis of Partitioning Algorithms for the Traveling Salesman Problem. *Math. Oper. Res.* **2** 209–224.

———. (1979). A Patching Algorithm for the Nonsymmetric Traveling-Salesman Problem. *SIAM J. Comput.* **8** 561–573.

———. (1980). An Algorithm to Solve the $m \times n$ Assignment Problem in Expected Time $O(mn \log n)$. *Networks* **10** 143–152.

Lawler, E. L. (1976). *Combinatorial Optimization: Networks and Matroids*. Holt, Reinhart and Winston, New York.

Lueker, G. S. (1981). Optimization Problems on Graphs with Independent Random Edge Weights. *SIAM J. Comput.* **10** 338–351.

Marchetti-Spaccamela, A. and Talamo, M. (1982). Probabilistic Analysis of Two Euclidean Location Problems. *RAIRO Inform. Théor.* (to appear).

Megiddo, N. (1979). Combinatorial Optimization with Rational Objective Functions. *Math. Oper. Res.* **4** 414–424.

———. (1981). Applying Parallel Computation Algorithms in the Design of Serial Algorithms. *Proc. IEEE 22nd Sympos. of Foundations of Computer Science*.

Papadimitriou, C. H. (1981). Worst Case and Probabilistic Analysis of a Geometric Location Problem. *SIAM J. Comput.* **10** (3) 542–557.

Rényi, A. (1970). *Foundations of Probability*. Holden Day, Inc., San Francisco.

Rohlf, F. G. (1978). A Probabilistic Minimum Spanning Tree Algorithm. *Inform. Process Lett.* **7** 44–48.

Spira, P. M. (1973). A New Algorithm for Finding all Shortest Paths in a Graph of Positive Arcs in Average Time $O(n^2 \log^2 n)$. *SIAM J. Comput.* **2** 28–32.

Steele, M. J. (1981). Complete Convergence of Short Paths and Karp's Algorithm for the T.S.P. *Math. Oper. Res.* **6** 374–378.

Walkup, D. (1979). On the Expected Value of a Random Assignment Problem. *SIAM J. Comput.* **8** 440–442.

Weide, B. W. (1980). Optimal Time Algorithms for Closest Point Problems. *ACM Trans Math. Software* **6** 563–580.

Yao, A. C. (1975). An $O(|E| \log \log |V|)$ Algorithm for Finding Minimum Spanning Trees. *Inform. Process Lett.* **4** 21–23.

Zemel, E. (1981). On Search Over Rationals. *Oper. Res. Lett.* **1** 34–38.

———. (1982). Probabilistic Analysis of Geometric Location Problems. *SIAM J. Algebraic Discrete Methods* (to appear).

HASSIN: DEPARTMENT OF STATISTICS, TEL AVIV UNIVERSITY, TEL AVIV 69978, ISRAEL

ZEMEL: J. L. KELLOGG GRADUATE SCHOOL OF MANAGEMENT, NORTHWESTERN UNIVERSITY, EVANSTON, ILLINOIS 60201

LEON RECAMATI GRADUATE SCHOOL OF BUSINESS ADMINISTRATION, TEL AVIV UNIVERSITY, TEL AVIV 69978, ISRAEL