OPTIMUM LOCATIONS ON A GRAPH WITH PROBABILISTIC DEMANDS

by

H. Frank

ERL Technical Memorandum M-131
27 October 1965

ELECTRONICS RESEARCH LABORATORY
University of California, Berkeley

Optimum Locations on a Graph with Probabilistic Demands*

H. Frank†
University of California, Berkeley

ABSTRACT

The traffic demands at the stations of a communication network are usually not deterministic. Optimum locations found using deterministic techniques are poor when the random nature of the network traffic is considered. The concepts of absolute centers and medians are generalized to maximum probability absolute centers and medians. Minimum variance points are also considered. Techniques to locate these optimum points are discussed.

INTRODUCTION

A communication or traffic network may usually be represented by a finite graph G, with weights attached to each of its branches and vertices. Given a graph G, one may be asked to find a point on G which is "optimum" in some predetermined sense. For example,

---

*This work was supported in part by the U. S. Army Research Office, Durham, N. C., Grant No. DA-ARO-D-31-124-G576.

†Department of Electrical Engineering, University of California, Berkeley, California; formerly with Department of Electrical Engineering, Northwestern University, Evanston, Illinois.

one may be asked to find a point on G from which the farthest point can be reached in the shortest time. In the simplest case, finding an optimum point may be equivalent to finding the center or median of G[1,2].

Hakimi considered the problem of finding the absolute centers and medians of a graph [3,4]. In the formulation of his model, the weight w_i of branch B_i $(i=1,\ldots,B)$ represents the length (or cost per unit length) of that branch, while the weight h_j of vertex v_j $(j=1,\ldots,n)$ is the average traffic (number of accidents or messages) occuring at v_j.

In reality, the traffic occuring at a vertex is not a fixed number, but rather a random number with a possibly known probability distribution. Consequently, an optimum point found by deterministic methods will vary with different realizations of the random events. It is therefore necessary to reevaluate the "optimality" of the deterministic optimum points. The concepts of absolute center and median must be generalized.

ABSOLUTE EXPECTED CENTERS AND MEDIANS

Consider a weighted n vertex graph G; a point X on G, is a point on some branch of G. The distance, d(X, Y), between any two points X and Y on G, is the length of the shortest path connecting X and Y (the length of a path is the sum of the branch weights in that path) [5]. Let H_i $(i=1,\ldots,n)$ be a nonnegative random variable corresponding to the weight of vertex v_i. With the above notions, we make the following definitions.[1]

<u>Definition 1.</u> A point X_{oe} on a branch of G is an <u>absolute expected center</u> (AEC) of G, if for every point X on G

$$\max_{1 \leq i \leq n} EH_i\, d(v_i, X_{oe}) \leq \max_{1 \leq i \leq n} EH_i\, d(v_i, X) \tag{1}$$

Definition 2. A point Y_{oe} on a branch of G is an absolute expected median (AEM) of G, if for every point Y on G

$$\sum_{i=1}^{n} EH_i\, d(v_i, Y_{oe}) \leq \sum_{i=1}^{n} EH_i\, d(v_i, Y) \tag{2}$$

Note that these definitions are identical to Hakimi's definitions in the case of deterministic vertex weights.

Definition 3. The expected radius r_o of G is a number defined by

$$r_o = \min_{X \text{ on } G} \max_{1 \leq i \leq n} EH_i\, d(v_i, X) = \max_{1 \leq i \leq n} EH_i\, d(v_i, X_{oe}) \tag{3}$$

Definition 4. The expected median length R_o of G is a number defined by

$$R_o = \min_{Y \text{ on } G} \sum_{i=1}^{n} EH_i\, d(v_i, Y) = \sum_{i=1}^{n} EH_i\, d(v_i, Y_{oe}) \tag{4}$$

Suppose that the AEC and AEM of G are found [3]. A natural question is: How "optimum" are these points when the random weights of G are considered. The interpretation of "optimum" for deterministic graphs is clear. The AEC will be no "farther" than r_o from any vertex of G; the median length of G will be no greater than R_o. If h_i is the

---

[1] E represents the expectation operator.

deterministic weight of v_i, equivalent statements for the above relationship are[2]

$$\text{Prob}\left\{ \max_{1 \leq i \leq n} h_i\, d(v_i, X_{oe}) > r \right\} = 0 \quad r \geq r_o \tag{5}$$

$$\text{Prob}\left\{ \sum_{i=1}^{n} h_i\, d(v_i, Y_{oe}) > R \right\} = 0 \quad R \geq R_o \tag{6}$$

If h_i is a random number, Eqs. (5) and (6) may not be true. It is thus desirable to investigate the quantities on the left hand sides of these equations. The following two theorems provide upper bounds for these numbers.

<u>Theorem 1</u>. Let X_{oe} be an absolute expected center of a graph G, whose vertices $v_1, \ldots, v_n$ are weighted with the nonnegative, independent random variables $H_1, \ldots, H_n$, respectively. Then

$$\text{Prob}\left\{ \max_{1 \leq i \leq n} H_i\, d(v_i, X) \geq r \right\} \leq \frac{1}{r} \sum_{i=1}^{n} EH_i\, d(v_i, X)$$

$$- \frac{1}{r^2} \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} EH_i\, EH_j\, d(v_i, X)\, d(v_j, X)$$

$$+ \ldots + \frac{(-1)^{n-1}}{r^n} EH_1 \ldots EH_n\, d(v_1, X) \ldots d(v_n, X) \tag{7}$$

---

2 If U is a random variable, Prob $\{ U > z \}$ represents the probability that U is greater than z.

and in particular

$$\text{Prob}\left\{\max_{1\leq i\leq n} H_i\, d(v_i, X_{oe}) \geq r\right\} \leq 1 - \left(1 - \frac{r_o}{r}\right)^n \tag{8}$$

Proof. Since $H_1, \ldots, H_n$ are independent,

$$\text{Prob}\left\{\max_{1\leq i\leq n} H_i\, d(v_i, X) \geq r\right\} = 1 - \prod_{i=1}^{n} \text{Prob}\left\{H_i\, d(v_i, X) < r\right\} \tag{9}$$

and if U is a nonnegative random variable, it is easy to show [6] that

$$\text{Prob}\left\{U < r\right\} \geq 1 - \frac{EU}{r} \tag{10}$$

Replacing $\text{Prob}\left\{H_i d(v_i, X) < r\right\}$ by $\left(1 - EH_i\, d(v_i, X)\right)$ in the right hand side of Eq. (9) gives

$$\text{Prob}\left\{\max_{1\leq i\leq n} H_i\, d(v_i, X) \geq r\right\} \leq 1 - \prod_{i=1}^{n} (1 - EH_i\, d(v_i, X)/r) \tag{11}$$

Expanding the right hand side of Eq. (11) gives the desired result of Eq. (7).

If we let $X = X_{oe}$, Eq. (11) gives

$$\text{Prob}\left\{\max_{1\leq i\leq n} H_i\, d(v_i, X_{oe}) \geq r\right\} \leq 1 - \prod_{i=1}^{n} (1 - EH_i\, d(v_i, X_{oe})/r) \tag{12}$$

Clearly,

$$\prod_{i=1}^{n} (1 - EH_i\, d(v_i, X_{oe})/r) \geq (1 - \max_{1 \leq i \leq n} EH_i\, d(v_i, X_{oe})/r)^n \tag{13}$$

and by the definition of r_o, Eqs. (12) and (13) imply

$$\text{Prob}\left\{ \max_{1 \leq i \leq n} EH_i\, d(v_i, X_{oe}) \geq r \right\} \leq 1 - (1 - \max_{1 \leq i \leq n} EH_i\, d(v_i, X_{oe})/r)^n$$

$$= 1 - (1 - r_o/r)^n \tag{14}$$

<u>Theorem 2.</u> Let Y_{oe} be an absolute expected median of a graph G, whose vertices are weighted with the nonnegative random variables $H_1, \ldots, H_n$. Then

$$\text{Prob}\left\{ \sum_{i=1}^{n} H_i\, d(v_i, Y_{oe}) \geq R \right\} \leq \frac{R_o}{R} \tag{15}$$

<u>Proof.</u> Using the same method as in Theorem 1, we have

$$\text{Prob}\left\{ \sum_{i=1}^{n} H_i\, d(v_i, Y) \geq R \right\} \leq (E \sum_{i=1}^{n} H_i\, d(v_i, Y))/R$$

$$= \frac{1}{R} \sum_{i=1}^{n} EH_i\, d(v_i, Y) \tag{16}$$

And if $Y = Y_{oe}$,

$$\text{Prob}\left\{ \sum_{i=1}^{n} H_i\, d(v_i, Y_{oe}) \geq R \right\} \leq \frac{1}{R} \sum_{i=1}^{n} EH_i\, d(v_i, Y_{oe}) = \frac{R_o}{R} \tag{17}$$

It is interesting to note that if we attempt to use the results of Theorems 1 and 2 to find points which minimize the upper bounds in Eqs. (7) or (16), we find that Y_{oe} minimizes the bound in Eq. (16) but that X_{oe} does not necessarily minimize the bound in Eq. (7). Additional bounds are given in the following section.

MAXIMUM PROBABILITY ABSOLUTE CENTERS AND MEDIANS

In the preceding section we obtained some simple upper bounds for $\text{Prob}\left\{\max H_i\, d(v_i, X) \geq r\right\}$ and $\text{Prob}\left\{\sum H_i\, d(v_i, Y) \geq R\right\}$.

Given a number r (or R), a natural goal is to find an X_o (or Y_o) on G, such that $\text{Prob}\left\{\max H_i\, d(v_i, X_o) \geq r\right\}$ (or $\text{Prob} \sum H_i\, d(v_i, Y_o) \geq R$) is minimized. In other words, we want to find a point X_o such that the greatest weighted distance stays within an allowable limit with maximum probability. These ideas lead to the following definitions.

<u>Definition 5.</u> A point X_o on a branch of G is a <u>maximum probability absolute r center</u> (MPArC) of G if for every point X on G

$$\text{Prob}\left\{\max_{1\leq i\leq n} H_i\, d(v_i, X_o) \geq r\right\} \leq \text{Prob}\left\{\max_{1\leq i\leq n} H_i\, d(v_i, X) \geq r\right\} \tag{18}$$

<u>Definition 6.</u> A point Y_o on a branch of G, is a <u>maximum probability absolute R median</u> (MPARM) of G, if for every point Y on G

$$\text{Prob}\left\{\sum_{i=1}^{n} H_i\, d(v_i, Y_o) \geq R\right\} \leq \text{Prob}\left\{\sum_{i=1}^{n} H_i\, d(v_i, Y) \geq R\right\} \tag{19}$$

The above definitions can be written in expanded form by introducing the concept of a local optimum point [3]. We demonstrate with

Definition 5. Let X_{oj} be a point on branch b_j such that

$$\text{Prob}\left\{\max_{1\le i\le n} H_i\, d(v_i, X_{oj}) \ge r\right\} = \min_{X \text{ on } b_j} \text{Prob}\left\{\max_{1\le i\le n} H_i\, d(v_i, X) \ge r\right\} \tag{20}$$

Then, X_{oj} is a local MPArC and X_o is an X_{oj} which minimizes the left hand side of Eq. (20). Consequently, if the set of local optimum points of G are available, the optimum point can be easily obtained.

Consider the graph of Figure 1. If X is an arbitrary point on branch b_k, then

$$d(v_i, X) = \min[x + d(v_p, v_i),\ \ B_k - x + d(v_q, v_i)] \tag{21}$$

and, it may be easily shown that

$$\text{Prob}\left\{H_i\, d(v_i, X) < r\right\} \triangleq \tilde{F}_i(x) = \begin{cases} F_i\left[\dfrac{r}{x + d(v_p, v_i)}\right] & 0 \le x \le a_i \\[2ex] F_i\left[\dfrac{r}{B_k - x + d(v_q, v_i)}\right] & a_i < x \le B_k \end{cases} \tag{22}$$

where $F_i[z]$ is the cumulative probability distribution function of $H_i\,(i=1,\ldots,n)$ and $a_i = \frac{1}{2}[B_k + d(v_q, v_i) - d(v_p, v_i)]$. The above relationship is illustrated in Figure 2.

From Figure 2 and Eq. (9), we can immediately obtain the following probability bounds for X_{ok}, a local MPArC.

$$\prod_{i=1}^{n} \min\left\{F_i\left[\frac{r}{d(v_p, v_i)}\right] \quad , \quad F_i\left[\frac{r}{d(v_q, v_i)}\right]\right\}$$

$$\leq \mathrm{Prob}\left\{\max_{1 \leq i \leq n} H_i\, d(v_i, X_{ok}) < r\right\}$$

$$\leq \prod_{i=1}^{n} \max\left\{F_i\left[\frac{r}{d(v_p, v_i)}\right] \quad , \quad F_i\left[\frac{r}{d(v_q, v_i)}\right]\right\} \tag{23}$$

The problem of finding a local MPArC is equivalent to finding an x which maximizes $\prod \tilde{F}_i(x)$ and is numerically straight forward. However, if the random variables $H_1, \ldots, H_n$ are discrete, the following interesting result is obtained.

<u>Theorem 3.</u> Let $H_1, \ldots, H_n$ be discrete independent random variables. Then, there exists an interval (possibly degenerate) $[X^*, X^{**}]$ on each branch of G, such that any $X \in [X^*, X^{**}]$ is a local maximum probability absolute r center of G.

<u>Proof.</u> Let $F_i[z]$ $(i=1, \ldots, n)$ be the cumulative probability distribution function of H_i such that

$$F_i[z] = k_{ij} \qquad j-1 \leq z < j \qquad j = 1, 2, \ldots \qquad k_{i_0} = 0 \tag{24}$$

The probability distribution of $d(v_i, X)\, H_i$ is $F_i\left[\frac{z}{d(v_i, X)}\right]$ and for $z = r$, is a function $\tilde{F}_i(x)$ shown in Figure 3 and given by Eq. (25a) for $0 \leq x \leq a_i$ and by Eq. (25b) for $a_i < x \leq B_k$.

$$\tilde{F}_i(x) = \begin{cases} k_{is} = F\left[\dfrac{r}{d(v_i, v_p)}\right] & 0 \leq x \leq \dfrac{r}{s-1} - d(v_i, v_p) \\[2ex] k_{i(s-j)} & \dfrac{r}{s-j} - d(v_i, v_p) < x \leq \dfrac{r}{s-j-1} - d(v_i, v_p) \\[2ex] 0 & j \geq s \end{cases}$$

(25 a)

$$\tilde{F}_i(x) = \begin{cases} k_{i(s-m)} = \tilde{F}_i(a_i) & a_i < x \leq B_k + d(v_i, v_q) - \dfrac{r}{s-m+1} \\[2ex] k_{i(s-m+j)} & B_k + d(v_i, v_q) - \dfrac{r}{s-m+j} \leq x < \\[2ex] & B_k + d(v_i, v_q) - \dfrac{r}{s-m+j+1} \end{cases}$$

(25 b)

Let $\Pi_i = \left\{ x_{i_0} = 0, x_{i1}, \ldots, x_{it_i} \right\}$ $(i = 1, \ldots, n)$ be the set of jump points of $\tilde{F}_i(x)$ and let $\Pi = \left\{ \omega_0 = 0, \omega_1, \ldots, \omega_t \right\}$ $(t \leq \sum t_i)$ be the common refinement of the Π_i. Clearly, $\tilde{F}_i(x)$ is constant for all i on the open interval (ω_{j-1}, ω_j) for $j = 1, \ldots, t$. Therefore, $\text{Prob}\left\{ \max H_i \, d(v_i, X) < r \right\}$ is also constant on each such interval. Then, either there exists at least one interval, say $(\omega_{\mu-1}, \omega_\mu)$ where this probability is maximum, or the maximum occurs only at an ω_j of Π.

Note that a maximum probability point must occur at an ω_j of Π, regardless of the existence of a maximum probability interval. Hence, local MPArC's could readily be found for a graph G by computing the jump points of the $\tilde{F}_i(x)$.

Remark. It is also possible to show that if the H_j are discrete, independent random variables, and there exists a non-degenerate, maximum probability interval (X^*, X^{**}) such that any X in that interval is a $MPAr_1C$, then there exists an $\epsilon > 0$, such that if $|r_1 - r_2| < \epsilon$, there exists an $X' \epsilon (X^*, X^{**})$ such that X' is a local $MPAr_2C$ of G.

We now turn our attention to the maximum probability absolute R medians of G. Consider the quantity $\sum_{i=1}^{n} H_i\, d(v_i, Y)$. If n is "reasonably" large and the H_i are independent, for each fixed Y, we can approximate the sum by a normal random variable Z with mean $\mu(Y)$ and variance $\sigma^2(Y)$ where

$$\mu(Y) = \sum_{i=1}^{n} EH_i\, d(v_i, Y) \tag{26}$$

$$\sigma^2(Y) = \sum_{i=1}^{n} \operatorname{Var} H_i\, d^2(v_i, Y) \tag{27}$$

Therefore,

$$\operatorname{Prob}\left\{ Z < R \right\} = \Phi\left(\frac{R - \mu(Y)}{\sigma(Y)} \right) \tag{28}$$

where Φ is the standard normal distribution function. Since Φ is a strictly increasing function, $Y = Y_o$ maximizes $\operatorname{Prob}\left\{ Z < R \right\}$ if and only if $Y = Y_o$ maximizes

$$\frac{R - \mu(Y)}{\sigma(Y)} = \frac{R - \sum_{i=1}^{n} EH_i\, d(v_i, Y)}{\left(\sum_{i=1}^{n} \operatorname{Var} H_i\, d^2(v_i, Y)\right)^{1/2}} \tag{29}$$

From Eq. (29), we can immediately evaluate the "optimality" of an sbsolute expected median Y_{oe}.

$$\operatorname{Prob}\left\{ \sum_{i=1}^{n} H_i\, d(v_i, Y_{oe}) \leq R_o \right\} = \frac{1}{2} \tag{30}$$

Consequently, Y_{oe} is a "poor" location for the median of G. Maximizing Eq. (29) is a relatively routine numerical problem, and will not be discussed further. Thus, if the normal approximation is used, MPARM's of G can be located without great difficulty. If exact expressions are required, n-fold convolutions are encountered and the solution is considerably more complicated.

MINIMUM VARIANCE ABSOLUTE MEDIANS

In finding an optimum location for a median of G, the designer may be satisfied in knowing that the probabilistic demands for the system stay close to a nominal value. Hence, it could be desirable to find a point Y_{ov} on G such that the variance of $\sum H_i\, d(v_i, Y_{ov})$ is minimum.

Let $H_1, \ldots, H_n$ be independent random variables with variances $\sigma_{v_1}^2, \ldots, \sigma_{v_n}^2$, respectively. For Y on branch b_k,

$$\operatorname{Var}\left\{\sum_{i=1}^{n} H_i\, d(v_i, Y)\right\} = \sum_{i=1}^{n} \sigma_{v_i}^2 \left\{\min[y + d(v_p, v_i),\ B_k - y + d(v_q, v_i)]\right\}^2 \tag{31}$$

Minimizing Eq. (31) is again numerically straight forward. However, if b_k is an isthmus[3] of G, we obtain the following result.

<u>Theorem 4.</u> Let b_k be an isthmus of G, which if deleted divides G into subgraphs G_1 and G_2. Then, there exists a Y_{ov} on b_k such that

$$\operatorname{Var}\left\{\sum_{i=1}^{n} H_i\, d(v_i, Y_{ov})\right\} = \min_{Y \text{ on } b_k} \operatorname{Var}\left\{\sum_{i=1}^{n} H_i\, d(v_i, Y)\right\} \tag{32}$$

and

$$d(v_p, Y_{ov}) = \begin{cases} 0 & D \leq 0 \\ D & 0 < D < B_k \\ B_k & B_k \leq D \end{cases} \tag{33}$$

where

$$D = \frac{\displaystyle\sum_{v_i \in G_1} \sigma_{v_i}^2\, d(v_p, v_i) - \sum_{v_j \in G_2} \sigma_{v_j}^2\, d(v_p, v_j)}{\displaystyle\sum_{i=1}^{n} \sigma_{v_i}^2}$$

<u>Proof.</u> Since b_k is an isthmus of G, there are no paths from any vertex of G_1 to any vertex of G_2, except through branch b_k. Let $v_p \in G_1$ and $v_q \in G_2$. Then, for $v_i \in G_1$, $v_j \in G_2$ and Y on b_k,

[3] Any branch of a connected graph G, which if deleted from G divides the graph into two components (subgraphs) is called an isthmus.

$$d(v_i, Y) = d(v_i, v_p) + d(v_p, Y) \tag{34 a}$$

$$d(v_j, Y) = d(v_j, v_q) + d(v_q, Y) \tag{34 b}$$

or, referring to Figure 1,

$$d(v_i, Y) = d(v_i, v_p) + y \tag{35 a}$$

$$d(v_j, Y) = d(v_j, v_q) + B_k - y = d(v_j, v_p) - y \tag{35 b}$$

Therefore, from Eqs. (31) and (35),

$$\begin{aligned} \mathrm{Var} \sum_{i=1}^{n} H_i\, d(v_i, Y) &= \sum_{v_i \in G_1} \sigma_{v_i}^2 (d(v_i, v_p) + y)^2 + \sum_{v_j \in G_2} \sigma_{v_j}^2 (d(v_j, v_p) - y)^2 \\ &= y^2 \sum_{i=1}^{n} \sigma_{v_i}^2 + 2y\left[\sum_{v_i \in G_1} \sigma_{v_i}^2\, d(v_i, v_p) - \sum_{v_j \in G_2} \sigma_{v_j}^2\, d(v_j, v_p)\right] \\ &\quad + \sum_{i=1}^{n} \sigma_{v_i}^2\, d^2(v_i, v_p) \end{aligned} \tag{36}$$

Taking the derivative with respect to y of both sides of Eq. (36) gives

$$\frac{d}{dy} \mathrm{Var} \sum_{i=1}^{n} H_i\, d(v_i, Y) = 2y \sum_{i=1}^{n} \sigma_{v_i}^2 + 2 \sum_{v_i \in G_1} \sigma_{v_i}^2\, d(v_i, v_p) - 2 \sum_{v_j \in G_2} \sigma_{v_j}^2\, d(v_j, v_p) \tag{37}$$

Consequently, $\mathrm{Var} \sum H_i\, d(v_i, Y)$ attains a minimum at $y = D$ where

$$D = \frac{\sum_{v_i \in G_1} \sigma_{v_i}^2 d(v_i, v_p) - \sum_{v_j \in G_2} \sigma_{v_i}^2 d(v_j, v_p)}{\sum_{i=1}^{n} \sigma_{v_i}^2} \tag{38}$$

and if $0 \leq d \leq B_k$, then Y_{ov}, located on b_k by Eq. (38), is a local minimum variance median. If $D < 0$, then Y_{ov} is at v_p and if $D > B_k$, then Y_{ov} is at v_q.

Corollary. If G is a tree, the minimum variance absolute median has a location on some branch b_k given by Eq. (33).

Proof. If G is a tree, every branch of G is an isthmus of G.

Generalization of Theorem 4. The proof of Theorem 4 uses the fact that if Y is on an isthmus of G, $d(v_i, Y) = d(v_i, v_s) + d(v_s, Y)$ where $s = p$ or $s = q$. Let G be a connected graph and b_k any branch of G. Then, from Eq. (21), $d(v_i, Y) = d(v_i, v_p) + y$ for $0 \leq y \leq a_i$ and $d(v_i, Y) = d(v_i, v_q) + B_k - y$ for $a_i < y \leq B_k$. Suppose we introduce a new vertex at $y = a_i$ and label the subdivided branch as b_{j1} and b_{j2}; for the purposes of computing $d(v_i, Y)$, Eqs. (34) hold on each branch segment. Let us generate all of the $a_i (i = 1, \ldots, n)$ and reorder them such that $a_i \leq a_{i+1}$. The a_i divide b_j into w segments $(w \leq n+1)$, and on each such segment Eqs. (34) hold. Consequently, the appropriate modification of Eq. (33) can be used to find the minimum variance point on each segment. The local minimum variance absolute median is easily found among the set of these points.

Remark. A natural, parallel approach to the above problem is to find a minimum variance absolute center, i.e., a point X on G which minimizes $\mathrm{Var}\{\max H_i\, d(v_i, X)\}$. This problem is far more complicated than the above because of the difficulty of obtaining analytic expressions for the moments of a finite set of non-identical random variables [7]. Even if normality assumptions are made, the approximate expressions encountered are complicated [7] and the problem is exceedingly cumbersome.

OPTIMUM LOCATIONS ON A GRAPH WITH SAMPLED POPULATIONS

Often, it is unreasonable to assume that the probability distributions of the H_i are known. A more reasonable assumption is that a set of observations $h_{i1}, h_{i2}, \ldots, h_{im_i}$ of H_i are available. Let the random variable H_{ij} $(i = 1, \ldots, n;\ j = 1, \ldots, m_i)$ correspond to the jth observation of H_i and suppose that $H_{i1}, \ldots, H_{im_i}$ are identically and independently distributed. Furthermore, we will assume that $F_i[z]$, the unknown probability distribution of H_i is strictly increasing; that is, H_i is a continuous random variable.

On the basis of the observed values of H_i, we would like to make a decision concerning the location of the optimum points of G. Let $S_{m_i}(z)$ be the empirical distribution function of H_i [6]. In other words,

$$S_{m_i}(z) = \frac{1}{m_i} \sum_{j=1}^{m_i} f_z(H_{ij}) \tag{39}$$

where

$$f_z(U) = \begin{cases} 0 & U \geq z \\ 1 & U < z \end{cases}$$

Thus, $m_i S_{m_i}(z)$ is the number of samples of H_i which are smaller than z. The following theorem shows that if the m_i are large, local maximum probability absolute centers found by using the S_{m_i} as the true distributions of the H_i will be "close" to the actual centers of G.

<u>Theorem 5</u>. Let $S_{m_i}(z)$ be the empirical distribution function of the continuous random variable H_i whose true distribution is $F_i[z](i=1,\ldots,n)$. Let $X_o(m_1,\ldots,m_n)$ be a local MPArC on branch b_k, found with respect to the S_{m_i}. Then, for $\delta > 0$, there exists an integer M, and a point X_o on b_k, such that for $m_i \geq M$ $(i=1,\ldots,n)$,

$$\text{Prob}\left\{ d(X_o, X_o(m_1,\ldots,m_n)) < \delta \right\} \to 1 \text{ as } M \to \infty \tag{40}$$

and X_o is a local maximum probability absolute r center of G.

<u>Proof</u>. (a) We first prove that for any $\epsilon > 0$, there exists an M such that for $m_i > M$ $(i=1,\ldots,n)$,

$$\text{Prob}\left\{ \left| \prod_{i=1}^{n} S_{m_i}(z) - \prod_{i=1}^{n} F_i[z] \right| < \epsilon \right\} \to 1 \text{ as } M \to \infty \tag{41}$$

By Glivenko's Theorem [6], for $\beta > 0$,

$$\text{Prob}\left\{ \sup_z \left| S_{m_i}(z) - F_i[z] \right| < \beta \right\} \to 1 \text{ as } m_i \to \infty \tag{42}$$

which implies

$$\text{Prob}\left\{ F_i[z] - \beta < S_{m_i}(z) < F_i[z] + \beta \right\} \rightarrow 1 \text{ as } m_i \rightarrow \infty \tag{43}$$

For M an integer and $m_i > M(i=1,\ldots,n)$ we have

$$\text{Prob}\left\{ \prod_{i=1}^{n} (F_i[z] - \beta) < \prod_{i=1}^{n} S_{m_i}(z) < \prod_{i=1}^{n} (F_i[z] + \beta) \right\} \rightarrow 1 \text{ as } M \rightarrow \infty \tag{44}$$

and since the inequalities

$$\prod_{i=1}^{n} (F_i[z] + \beta) \leq \prod_{i=1}^{n} F_i[z] - 1 + (1+\beta)^n = \prod_{i=1}^{n} F_i[z] + \sum_{k=1}^{n} \beta^k \tag{45 a}$$

and

$$\prod_{i=1}^{n} (F_i[z] - \beta) \geq \prod_{i=1}^{n} F_i[z] + (-\beta)^n \geq \prod_{i=1}^{n} F_i[z] - \sum_{k=1}^{n} \beta^k \tag{45 b}$$

hold, then

$$\text{Prob}\left\{ -\sum_{k=1}^{n} \beta^k < \prod_{i=1}^{n} S_{m_i}(z) - \prod_{i=1}^{n} F_i[z] < \sum_{k=1}^{n} \beta^k \right\} \rightarrow 1 \text{ as } M \rightarrow \infty \tag{46}$$

If $\beta = (\epsilon + 1)^{1/n} - 1$, Eq. (41) follows.

(b) Let $X = X_o$ maximize

$$F(X) \triangleq \prod_{i=1}^{n} F_i\left[\frac{r}{d(v_i, X)}\right] \tag{47 a}$$

and, let $X = X_{om} \triangleq X_o(m_1, \ldots, m_n)$ maximize the random variable

$$S(X) \triangleq \prod_{i=1} s_{m_i}\left(\frac{r}{d(v_i, X)}\right) \tag{47 b}$$

Then, we will prove that for $m_i > M$

$$\text{Prob}\left\{ F(X_{om}) > F(X_o) - 2\epsilon \right\} \to 1 \text{ as } M \to \infty \tag{48}$$

From Eq. (41)

$$\text{Prob}\left\{ F(X_o) - \epsilon < S(X_o) < F(X_o) + \epsilon \right\} \to 1 \tag{49}$$

and if

$$F(X_{om}) = F(X_o) - \gamma \qquad (\gamma > 0) \tag{50}$$

it follows that

$$\text{Prob}\left\{ F(X_o) - \epsilon - \gamma < S(X_{om}) < F(X_o) - \gamma + \epsilon \right\} \to 1 \tag{51}$$

However, if $\gamma > 2\epsilon$, we obtain

$$\text{Prob}\left\{ S(X_{om}) < F(X_o) - \epsilon < S(X_o) \right\} \to 1 \tag{52}$$

This is impossible since Eq. (47 b), $S(X_{om}) \geq S(X_o)$.

(c) Clearly, around each optimum point X_o on b_k, Eq. (49) defines an allowable region for X_{om}. Let x_o and x_{om} be the locations of X_o and X_{om} on b_k, respectively. Within a sufficiently small ϵ region, $F(X)$ must be strictly increasing for $x < x_o$ and strictly decreasing for $x > x_o$ (without loss of generality, we have assumed $F(X)$

is not flat at X_o). Consequently, within this region, the inverse function of F(X) exists and is continuous. Then, for any $\delta > 0$, we can find an $\epsilon > 0$ such that

$$|F(X_o) - F(X_{om})| < \epsilon \qquad \text{implies}$$

$$|x_o - x_{om}| < \delta \tag{53}$$

Since $|x_o - x_{om}| = d(X_o, X_{om})$ (for sufficiently small δ), the theorem is proved.

The problem of finding the maximum probability absolute medians of G on the basis of the population samples requires, in general, an n-fold convolution of the empirical distributions. However, the problem may be considerably simplified if we make the assumption that $\sum H_i d(v_i, Y)$ is a normal random variable. In this case, Eq. (29) must be maximized.

Let $\overline{H}_{m_i}$ and $V_{m_i}^2$ be defined by

$$\overline{H}_{m_i} \triangleq \frac{1}{m_i} \sum_{j=1}^{m_i} H_{ij} \qquad (i=1,\ldots,n) \tag{54a}$$

$$V_{m_i}^2 \triangleq \frac{1}{m_i} \sum_{j=1}^{m_i} (H_{ij} - \overline{H}_{m_i})^2 \qquad (i=1,\ldots,n) \tag{54b}$$

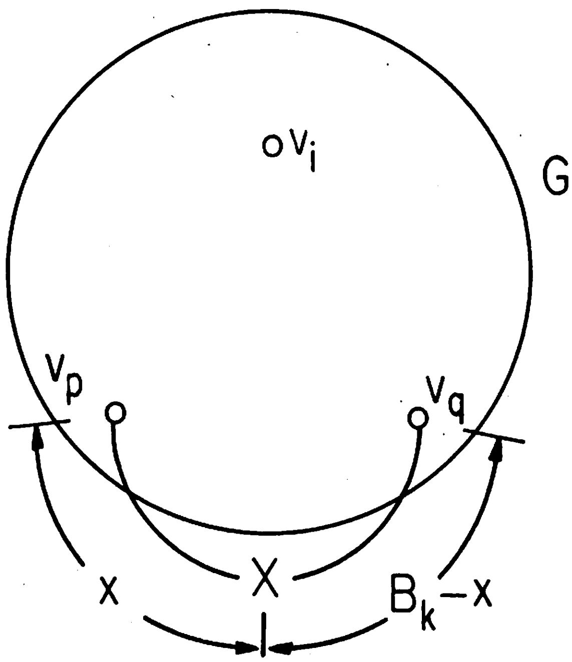
$\overline{H}_{m_i}$ and $V_{m_i}^2$ are consistent estimates of EH_i and $\sigma_{v_i}^2$, respectively (many other "nice" properties also hold)[6]. Consequently, it is not difficult to show that if $m_i > M\,(i=1,\ldots,n)$

$$\lim_{M\to\infty} \text{Prob}\left\{ \frac{R - \sum \bar{H}_{m_i} d(v_i, Y)}{\left(\sum V_{m_i}{}^2 d^2(v_i, Y)\right)^{1/2}} - \frac{R - \sum EH_i\, d(v_i, Y)}{\left(\sum \sigma_{v_i}{}^2 d^2(v_i, Y)\right)^{1/2}} > \epsilon \right\} = 0 \tag{55}$$

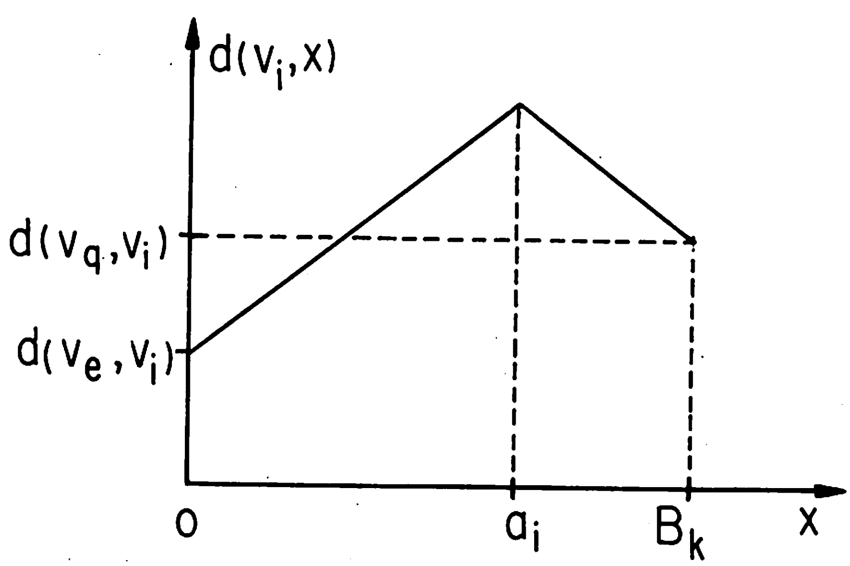
In other words, we may base our optimization procedure on the sample means and sample variances of the vertex populations.

CONCLUSIONS

The concepts of absolute median and absolute center were generalized to include the randomness of the network traffic. A number of problems remain unsolved. For example, if the random variables which represent the vertex demands are not independent, the location of maximum probability points on G becomes considerably more challenging. Hakimi generalized the absolute median to an absolute p-median [4]. A similar generalization can be made in the probabilistic case. The problem becomes much more complicated if distance is considered to be a random variable. The author is presently studying the feasibility of such an extension.



(a)



(b)

Fig. 1. Distance relations on branch b_k of G.

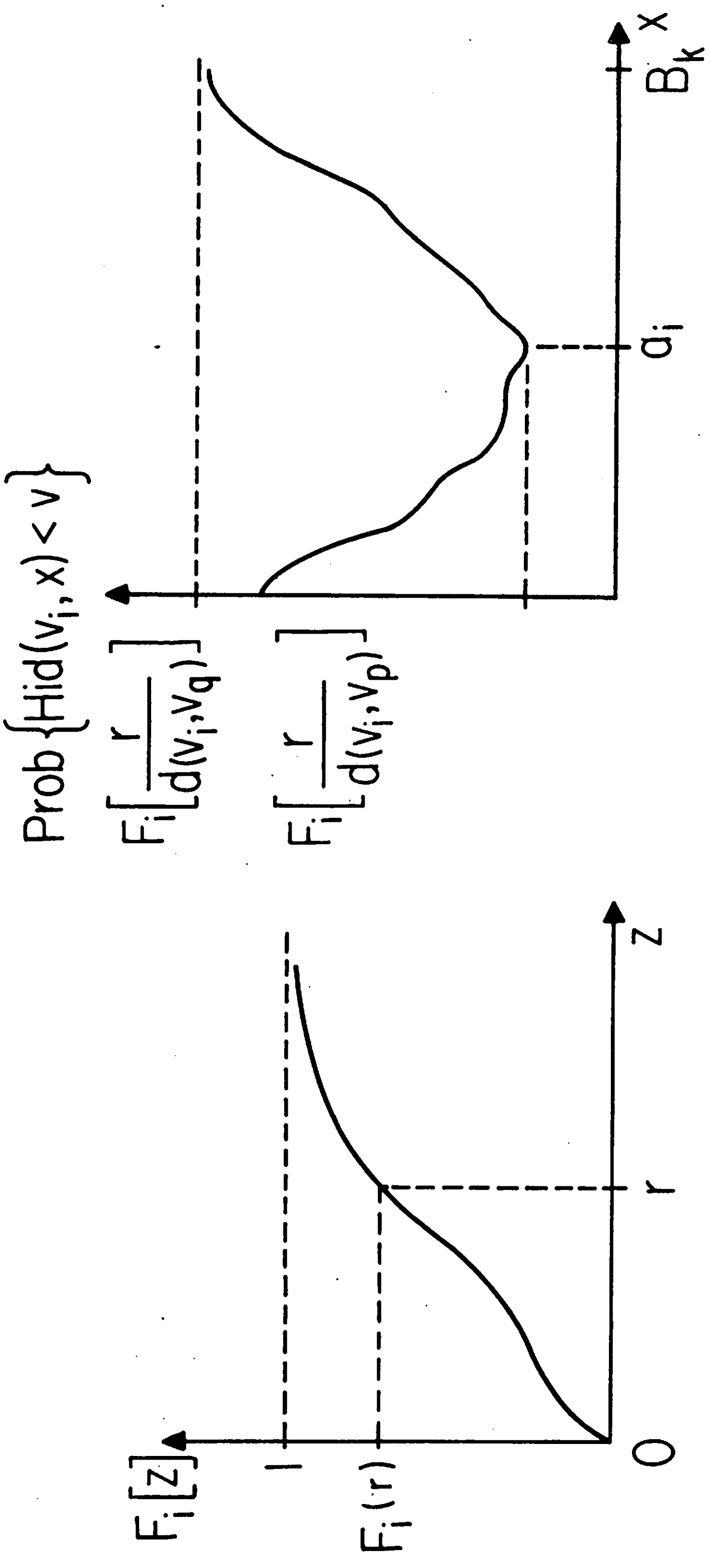


Fig. 2. Probability relations on branch b_k of G.

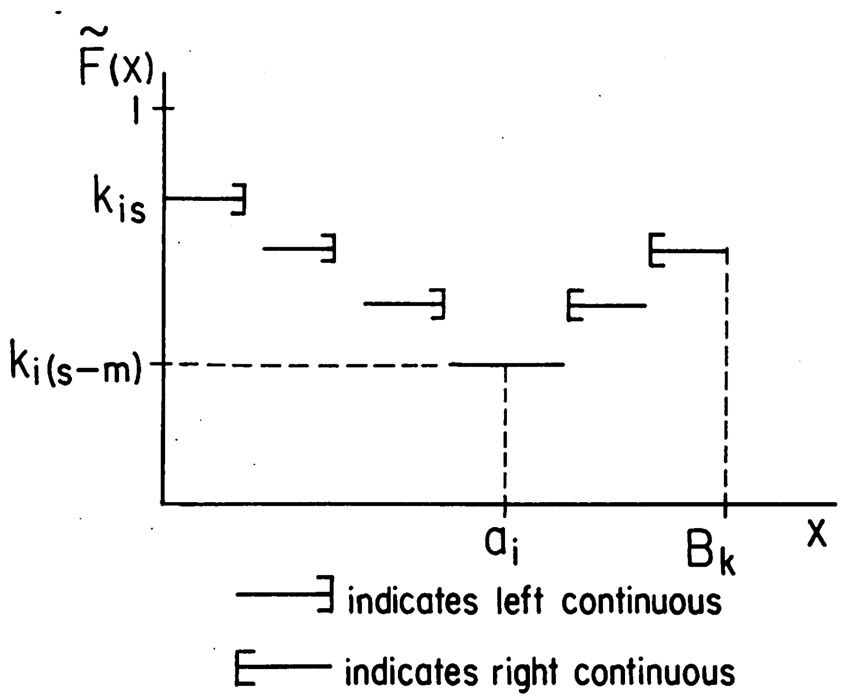


Fig. 3. Discrete probability relation on branch b_k of G.

REFERENCES

1. C. Berge, The Theory of Graphs, Methuen and Co., London; 1962.
2. O. Ore, "Theory of Graphs," Am. Math. Soc. Colloquium Publication, Providence, R. I.; 1962.
3. S. L. Hakimi, "Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph," J. Operations Research, Vol. 12, No. 3, pp. 450-459; May-June, 1964.
4. S. L. Hakimi, "Optimum Distribution of Switching Centers in a Communication Network and Some Related Problems," J. Operations Research, Vol. 13, No. 3, pp. 462-475; May-June 1965.
5. S. L. Hakimi and S. S. Yau, "Distance Matrix and the Synthesis of N Port Resistive Networks," Quarterly Applied Math, Vol. 22, No. 4, pp. 305-317; Jan. 1965.
6. M. Fisz, Probability Theory and Mathematical Statistics, John Wiley & Sons, Inc., New York; 1963.
7. C. E. Clark, "The Greatest of a Finite Set of Random Variables," J. Operations Research, Vol. 3, No. 2; March-April, 1961.