The Identity of Centroids and Medians in Discrete Sets

在離散集合中質量中心與中央點的一致性

Chih–Wei Chang[1] 張志維
Yue–Li Wang[1] 王有禮
Hon–Chan Chen[2] 陳宏昌

[1]Department of Information Management,
National Taiwan University of Science and Technology, Taipei
Email: ylwang@cs.ntust.edu.tw

[2]Department of Shipping Transportation and Management,
National Kaohsiung Institute of Marine Technology, Kaohsiung

Abstract

Centroids and medians are very important in discussing location problems. There are very few papers proposed for finding centroids and medians on various graphs. In some graphs, centroids and medians are the same. It was shown that the centroids are also the medians in a tree graph. In this paper, we shall prove the identity of centriods and medians in a discrete set according to the Manhattan metric on Z^2.

Keyword: Manhattan metric, centroids, medians.

1 Introduction

Location on networks is a topic of great importance in fields such as transportation, communication, service areas and computer sciences. Historically, the center and median have been proved useful as solutions for the locations of emergency and service facilities, respectively, such as a hospital, a police station, a post office, shopping mall, bank or power station. It was for telecommunication networks that Hakimi originally proposed median and center objectives for location of switching centers[5]. In [12], Slater introduced the competitive facility location problem in graphs. For example, each vertex in a graph G represents a customer and a store can be created in any vertex of G. Assume that every customer will shop at the closest store and that there is only one store in G before you want to create a new one to compete with it. Which vertex will be chosen as the location of your store so that you will have as many customers as possible? The solution of this problem is called a centroid.

Location problems in graphs and networks are studied wildly in operations research[3, 4, 8, 10, 12, 13, 14]. Most of researchers are focus on designing efficient algorithms for finding the centers, medians, and centroids of a graph. See [14] for a summary of early history. Very few papers discussed the identity of centers, medians, and centroids. However, Slater proved that, in a tree graph, a point u is a centroid if and only if u is also a median[12]. Moreover, he also proved that for any connected graph G, the centroids and medians are in the same block of G[13]. In this paper, we shall show the identity of medians and centroids of a discrete set according to the Manhattan metric on Z^2. A *discrete set* S is a subset of Z^2 with elements called *points*. In [9], Lungo et al. proposed an efficient algorithm for finding the medians of a discrete set S in $O(|R(S)|)$ time, where $R(S)$ denotes the smallest rectangle area containing S and $|R(S)|$ is the cardinality of $R(S)$. Chung extended their result to the weighted case and solved the problem in $O(|S|)$ time [2].

The rest part of this paper is organized as follows. In Section 2, we define our problem formally and introduce some notation in detail. Section 3 contains our main result of proving the identity of medians and centroids on a discrete set. Finally, concluding remarks are given in Section 4.

2 Preliminaries

We will follow the notation defined in [9]. For clarity, we introduce some of them as follows. Let $R(S)$ denote the smallest $m \times n$ rectangle containing S which is a discrete subset on Z^2. The position of each point p in $R(S)$ is indexed by (x_p, y_p), $1 \leq x_p \leq m$ and $1 \leq y_p \leq n$, where x_p and y_p are the row number and the column number, respectively. Note that we number the rows and columns starting from the lower-left corner of $R(S)$. Using the Manhattan metric, the *distance* between two points $p = (x_p, y_p)$ and $q = (x_q, y_q)$ is $d(p,q) = |x_p - x_q| + |y_p - y_q|$. The *total distance* of point p in $R(S)$ is $D(p) = \sum_{q \in S} d(p,q)$[1]. A point p in $R(S)$ is said to be a *median* with respect to S if $D(p)$ is minimum among the total distances of all points in $R(S)$. Generally, there are more than one median and some medians are in $R(S) - S$. We use $M(S)$ to denote the set of medians with respect to S. For a pair of points u, $v \in R(S)$, let P_{uv} be the set of points in S which are closer to u than v, including u itself if $u \in S$. Let $f(u,v) = |P_{uv}| - |P_{vu}|$ and $g(u) = min\{f(u,v)|v \in R(S) - u\}$. The *centroid value* of S, denoted by $c(S)$, is defined as $c(S) = max\{g(u)|u \in R(S)\}$. Point u is called a *centroid* if $g(u) = c(S)$. The set of centroids is denoted by $C(S)$.

We use an example to illustrate the above notation. In Figure 1, $R(S)$ is a 3×4 rectangle of 12 points, where S contains the points p_1, p_4, p_6, p_8, p_{10}, and p_{12}. The distances between point p_1 and points p_1, p_4, p_6, p_8, p_{10}, and p_{12} are 0, 3, 2, 4, 3, and 5, respectively. The total distance $D(p_1) = 0 + 3 + 2 + 4 + 3 + 5 = 17$. Similarly, the distances between point p_2 and points p_1, p_4, p_6, p_8, p_{10}, and p_{12} are 1, 2, 1, 3, 2, and 4, respectively, and $D(p_2) = 1 + 2 + 1 + 3 + 2 + 4 = 13$. For all points $p_i, i = 3, 4, \cdots, 12$, the values of $D(p_i)$ are 13, 13, 15, 11, 11, 11, 17, 13, 13, and 13, respectively. Therefore, the median set $M(S) = \{p_6, p_7, p_8\}$. Note that p_7 is not in S. Points p_6, p_8, p_{10}, and p_{12} are closer to p_6 than p_1, thus $|P_{p_6 p_1}| = 4$. Since no point is closer to p_1 than p_6 except p_1 itself, $|P_{p_1 p_6}| = 1$. Therefore, $f(p_6, p_1) = |P_{p_6 p_1}| - |P_{p_1 p_6}| = 4 - 1 = 3$. Similarly, $f(p_6, p_2) = 2$, $f(p_6, p_3) = 1$, $f(p_6, p_4) = 0$, $f(p_6, p_5) = 4$, $f(p_6, p_7) = 0$, $f(p_6, p_8) = 0$, $f(p_6, p_9) = 2$, $f(p_6, p_{10}) = 2$, $f(p_6, p_{11}) = 0$, and $f(p_6, p_{12}) = 0$. The value of $g(p_6)$ is 0. Using the same computation for each point in $R(S)$, the values of $g(p_i)$, $i = 1, 2, \cdots, 12$, are -4, -2, -1, -2, -4, 0, 0, 0, -4, -2, -2, and -2, respectively. By the definition of centroids, $C(S)$ contains points p_6, p_7, and p_8 which are also the medians with respect to S.

The ith *row projection* (respectively, jth *column projection*) of S is the number of points of S in the ith row (respectively, jth column). We denote the ith row projection by h_i, $1 \leq i \leq m$ and the jth *column projection* by v_j, $1 \leq j \leq n$. The vectors $H = (h_1, h_2, \cdots, h_m)$ and $V = (v_1, v_2, \cdots, v_n)$ are called the *horizontal* and *vertical projections*, respectively, of S. The prefix sums of vectors H and

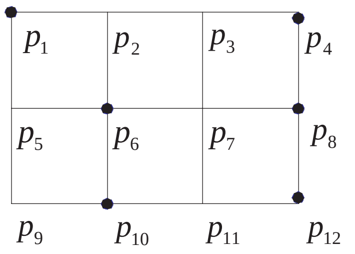

Figure 1: The Manhattan metric.

V are defined as follows:

$$H_0 = 0 \textit{ and } H_k = \sum_{i=1}^{k} h_i, k = 1, 2, ..., m$$

$$V_0 = 0 \textit{ and } V_k = \sum_{j=1}^{k} v_j, k = 1, 2, ..., n$$

$$\textit{and } A = \sum_{i=1}^{m} h_i = \sum_{j=1}^{n} v_j.$$

The ith row (respectively, jth column) of $R(S)$ is called a *median row* (respectively, *median column*) if $H_{i-1} \leq \frac{A}{2} \leq H_i$ (respectively, $V_{j-1} \leq \frac{A}{2} \leq V_j$). In [9], Lungo et al. proved that a point m belongs to $M(S)$ if and only if m is an intersection of a median row and a median column of $R(S)$. For example, in Figure 1, $H = (h_1, h_2, h_3) = (2, 2, 2)$, $V = (v_1, v_2, v_3, v_4) = (1, 2, 0, 3)$, $(H_1, H_2, H_3) = (2, 4, 6)$, $(V_1, V_2, V_3, V_4) = (1, 3, 3, 6)$ and the value of A is 6. It is easy to find that row 2 and columns 2, 3, and 4 are the median row and the median columns, respectively. The medians of $R(S)$, p_6, p_7, and p_8 are indeed the intersections of row 2 and columns 2, 3, and 4, respectively.

According to the result that a point m belongs to $M(S)$ if and only if m is an intersection of a median row and a median column of $R(S)$, the medians of S will form a rectangle. We denote it by $R(M)$. With respect to $R(M)$, we have the following observations:

Observation 1 *If m is a median and $m \in M(S)$, then m must be at one of the four corners of $R(M)$.*

Observation 2 *If there are more than two median rows (respectively, columns), then no point of S is in the intermediate median rows (respectively, columns).*

Observation 3 *If there are more than one median column (respectively, row) and m is a median at one of the left (respectively, lower) corners of $R(M)$, then $V_{y_m} = \frac{A}{2}$ (respectively, $H_{x_m} = \frac{A}{2}$). Furthermore, if m is a median at one of the right (respectively, upper) corners of $R(M)$, then $V_{y_m - 1} = \frac{A}{2}$ (respectively, $H_{x_m - 1} = \frac{A}{2}$) and $V_{y_m} > \frac{A}{2}$ (respectively, $H_{x_m} > \frac{A}{2}$).*

3 Main Result

In this section, we shall prove the identity of medians and centroids on a discrete set under the Manhattan metric. For ease of description, in Lemmas 1 and 2, all positions of the discrete set S are mapped into the coordinate system and a specific median is mapped to the origin. That is, if $m = (x_m, y_m)$ is the specific median, then for any point $u = (x_u, y_u)$, the abscissa and ordinate of u in the coordinate system are $y_u - y_m$ and $x_u - x_m$, respectively. Notice that the row number x_u is mapped to an ordinate and the column number y_u is mapped to an abscissa. We use $\hat{x}_u$ and $\hat{y}_u$ to denote the mapped abscissa and ordinate, respectively, of point u in the coordinate system. That is, $\hat{x}_u = y_u - y_m$ and $\hat{y}_u = x_u - x_m$.

Lemma 1 *Let $m = (x_m, y_m)$ and $u = (x_u, y_u)$ be two points in $R(S)$, where $m \in M(S)$ and $u \in R(S) - M(S)$. If $x_m = x_u$ or $y_m = y_u$, then $f(m, u) \geq 0$. Moreover, if m is the closest median to u, then $f(m, u) > 0$.*

Proof: Since m is at the origin, the coordinate of u is either $(0, k)$ or $(k, 0)$, where k is an integer. We only consider the case where $(\hat{x}_u, \hat{y}_u) = (k, 0)$ and

$k > 0$. The other cases can be handled similarly. Let $v = (x_v, y_v)$ be a point in $R(S)$ with $\hat{x}_v \leq 0$. It can be obtained that $d(m, v) < d(u, v)$ by the following derivation:

$$
\begin{aligned}
& d(m,v) \\
= \; & |0 - \hat{x}_v| + |0 - \hat{y}_v| \\
= \; & |\hat{x}_v| + |\hat{y}_v| \\
< \; & |k - \hat{x}_v| + |0 - \hat{y}_v| \\
= \; & d(u,v).
\end{aligned}
$$

Since $V_{y_m} \geq \frac{A}{2}$, there are at least $\frac{A}{2}$ points in S whose column numbers are less than or equal to y_m. Thus, there are at least $\frac{A}{2}$ points in S which are closer to m than u. It implies that $f(m, u) \geq 0$.

Now we prove that if m is the closest median to u, then $f(m, u) > 0$. Let m be the closest median to u. If $V_{y_m} = \frac{A}{2}$, then column $y_m + 1$ will be a median column and point $(x_m, y_m + 1)$ is also a median which is closer to u. It contradicts that m is the closest median to u. Therefore, $f(m, u) > 0$.

Q. E. D.

Let N_i contain the points of S in Quadrant i, $i = I, II, III, IV$. Note that the points in x-axis or y-axis are not in any quadrant. Let O be the origin and u be a point in S with $(\hat{x}_u, \hat{y}_u) = (k, k)$, where k is a positive integer. The Voronoi diagram[7, 11] of points u and O, under Manhattan metric, separates the plane into three regions: region A_u contains the points of S closer to u than O; region A_m contains the points of S closer to O than u; the third region contains the points of S which have the equal distance to u and O[7]. The points of S on x-axis and y-axis are separated into four segments L_0, R_0, D_0, and U_0: L_0 contains the points u of S with $\hat{x}_u \leq 0$ and $\hat{y}_u = 0$; R_0 contains the points u of S with $\hat{x}_u > 0$ and $\hat{y}_u = 0$; D_0 contains the points of S with $\hat{x}_u = 0$ and $\hat{y}_u < 0$; and U_0 contains the points of S with $\hat{x}_u = 0$ and $\hat{y}_u > 0$. See Figure 2 for an illustration.

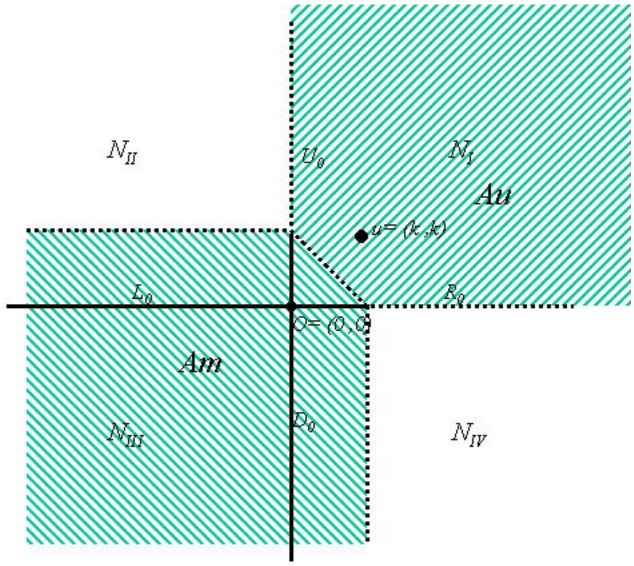

Figure 2: A Voronoi diagram under the Manhattan metric.

Lemma 2 *Let $m = (x_m, y_m)$ and $u = (x_u, y_u)$ be two points in $R(S)$, where $m \in M(S)$ and $u \in R(S) - M(S)$. If $x_m \neq x_u$ and $y_m \neq y_u$, then $f(m, u) \geq 0$. Moreover, if m is the closest median to u, then $f(m, u) > 0$.*

Proof: We only consider the case where point u is in Quadrant I. With a similar reasoning, we can prove the other cases in which the points are in other quadrants. Consider the following three cases.

Case 1. $\hat{x}_u > \hat{y}_u$.

In this case, we shall prove that $d(m, v) < d(u, v)$ for any point $v = (x_v, y_v)$ in S with $\hat{x}_v \leq 0$. Let w be a point in $R(S)$ whose coordinate is $(0, \hat{y}_u)$. Then,

$$
\begin{aligned}
& d(u,v) \\
= \; & d(u,w) + d(w,v) \\
= \; & |\hat{x}_u - 0| + |\hat{y}_u - \hat{y}_u| + d(w,v) \\
= \; & |\hat{x}_u| + d(w,v) \\
> \; & |\hat{y}_u| + d(w,v)
\end{aligned}
$$

$$\begin{aligned} &= d(m,w)+d(w,v) \\ &\geq d(m,v). \end{aligned}$$

Since $V_{y_m} \geq \frac{A}{2}$, there are at least $\frac{A}{2}$ points in S whose column numbers are less than or equal to y_m. Thus, there are at least $\frac{A}{2}$ points in S which are closer to m than u. It implies that $f(m,u) \geq 0$.

Case 2. $\hat{x}_u < \hat{y}_u$.

Similar to Case 1, it can be proved that there are at least $\frac{A}{2}$ points in S whose row numbers are less than or equal to x_m and $f(m,u) \geq 0$.

Case 3. $\hat{x}_u = \hat{y}_u$.

For the purpose of contradiction, we assume that $f(m,u) < 0$. That is $|P_{um}| > |P_{mu}|$ and $|A_u| > |A_m|$. Since A_u is a subset of N_I and A_m contains $N_{III} \bigcup L_0 \bigcup D_0$, $|N_I| \geq |A_u| > |A_m| = |N_{III}| + |L_0| + |D_0|$. Therefore,

$$\begin{aligned} |N_I| &> |N_{III}|+|L_0|+|D_0| \\ |N_I|+|N_{IV}|+|R_0| &> |N_{III}|+|L_0|+|D_0| \\ &\quad +|N_{IV}|+|R_0| \\ |N_I|+|N_{IV}|+|R_0| &> |N_{III}|+|N_{IV}|+h_{x_m} \\ &\quad +|D_0| \\ |N_I|+|N_{IV}|+|R_0| &> \frac{A}{2}. \end{aligned}$$

It implies that $V_{y_m} < \frac{A}{2}$ and contradicts that m is a median. Thus, $f(m,u) \geq 0$.

For the case where m is the closest median to u, we assume to the contrary that $f(m,u) \leq 0$. That is $|P_{um}| \geq |P_{mu}|$ and $|A_u| \geq |A_m|$. Since A_u is a subset of N_I and A_m contains $N_{III} \bigcup L_0 \bigcup D_0$, $|N_I| \geq |A_u| \geq |A_m| = |N_{III}| + |L_0| + |D_0|$. Moreover, m must be the upper-right corners of $R(M)$ since m is the closest median to u and u is in Quadrant I with $x_m \neq x_u$ and $y_m \neq y_u$. By Observation 3, $H_{x_m} = |N_{III}|+|N_{IV}|+h_{x_m}+|D_0| > \frac{A}{2}$. Therefore,

$$\begin{aligned} |N_I| &\geq |N_{III}|+|L_0|+|D_0| \\ |N_I|+|N_{IV}|+|R_0| &\geq |N_{III}|+|L_0|+|D_0| \\ &\quad +|N_{IV}|+|R_0| \\ |N_I|+|N_{IV}|+|R_0| &\geq |N_{III}|+|N_{IV}|+h_{x_m} \\ &\quad +|D_0| \\ |N_I|+|N_{IV}|+|R_0| &> \frac{A}{2}. \end{aligned}$$

It implies that $V_{y_m} < \frac{A}{2}$ and contradicts that m is a median. Thus, $f(m,u) > 0$.

Q. E. D.

Corollary 3 *If $u \in R(S)-M(S)$, then there exists a point w in $M(S)$ such that $f(w,u) > 0$.*

Lemma 4 *If $|M(S)| > 1$ and m, u belong to $M(S)$, then $f(m,u) = 0$.*

Proof: By Observations 1 and 2, the positions of m and u in $R(S)$ must be in one of the following cases:

1. m and u are at the same row,
2. m and u are at the same column, and
3. m and u are at the opposite corners of $R(M)$.

Suppose that m and u are at the same row, where $y_m < y_u$. By Observation 3, there are $\frac{A}{2}$ points closer to m than u since $V_{y_m} = \frac{A}{2}$. Moreover, since $V_{y_{u-1}} = \frac{A}{2}$, the number of points on column y_u and on the right hand side of u is $\frac{A}{2}$; that is, the number of points closer to u than m is $\frac{A}{2}$. We have $|P_{mu}| = |P_{um}| = \frac{A}{2}$ and $f(m,u) = 0$. With a similar reasoning, the same result for the latter two cases can be obtained.

Q. E. D

Lemma 5 *If point m is a median of S, then m is a centroid of S.*

Proof: By Lemmas 1 and 2 and the definition of $g(u) = min\{f(u,v)|v \in R(S)-u\}$, if u is a median, then $g(u) \geq 0$; otherwise, $g(u) \leq 0$. The definition of a centroid u is $g(u) = c(S) = max\{g(v)|v \in R(S)\}$. By Lemma 4, therefore, a median must be a centroid.

Q. E. D

Lemma 6 *If $u = (x_u, y_u) \in G$ is a centroid under the Manhattan metric, then u is a median.*

Proof: Assume to the contrary that u is a centroid but not a median. Let point m be the closest median to u. By Lemmas 1 and 2 and Corollary 3, we know that $g(m) \geq 0$ and $g(u) < 0$. It contradicts the definition of a centroid $g(u) = max\{g(v)|v \in G\}$. It completes the proof.

Q. E. D.

We summarize Lemmas 5 and 6 as the following theorem.

Theorem 7 *Under the Manhattan metric, point u is a centroid of G if and only if it is a median of G.*

4 Concluding Remarks

In this paper, we prove the identity of medians and centroids of a discrete set under Manhattan metric. Thus, an algorithm which finds the medians of a graph can also be applied to find the centroids of the graph and vice versa. By using the algorithm proposed by Chung[2], the centroids of a discrete set can be found in $O(|S|)$ time. In general graphs, a median may not be a centroid. It is worth to study the identity of medians and centroids on other graphs.

References

[1] F. Buckley and F. Harary, *Distance in Graphs*, Addison-Wesley, Reading, MA, 1990.

[2] Kuo-Liang Chung, On Finding Medians of Weighted Discrete Points, *Information Processing Letter*, Vol. 74, 2000, pp. 103-106.

[3] A. J. Goldman, Optimal Center Location in Simple Networks, *Transportation Science* Vol. 5, 1971, pp. 212-221.

[4] A. J. Goldman, Minimax Location in a Facility in a Network, *Transportation Science* Vol. 6, 1972, pp. 407-418.

[5] S.L. Hakimi, Optimum Location of Switching Centers and the Absolute Centers and Medians of a graph, *Operations Research*, Vol. 12, 1964, pp. 450-459.

[6] F. Harary, Status and Contrastatus, *Sociometry*, Vol. 22, 1959, pp. 23-430.

[7] R. Klein and A. Lingas, Manhattonian Proximity in a Simple Polygon, *Eighth Annual Symposium on Computational Geometry*, 1992, pp. 312-319.

[8] R. Laskar and D. Shier, On Powers and Centers of Chordal Graphs, *Discrete Apply Mathematics*, Vol. 6, 1983, pp. 139-147.

[9] A. Del Lungo, M. Nivat, R. Pinzani and L. Sorri, The Medians of Discrete Sets, *Information Processing Letter*, **65**, pp. 293-299, 1998.

[10] O. R. Oellermann, On Steiner Centers and Steiner Medians of Graphs Telecommunication Networks, *Networks*, Vol. 34, 1999, pp. 258-263.

[11] F. P. Preparata and M. I. Shamos, *Computational Geometry*, Springer-Verlag, New York, 1985.

[12] P. J. Slater, Maximum Facility Location,*Journal of Research of the National Bureau of Standards*, Vol. 79, 1975, pp. 107-115.

[13] P. J. Slater, Center, Median and Centroid Subgraphs, *Networks*, Vol. 34, 1999, pp. 303-311.

[14] B. C. Tansel, R. L. Francis and T. J. Lowe, Locations on Networks: A Survey, *Management Science*, Vol. 29, 1983, pp. 482-511.