

No. 836

Size-Density Hypothesis in p -Median Problems

by

Tsutomu Suzuki

October 1999

Size-Density Hypothesis in p -Median Problems

by

Tsutomu Suzuki

Assistant Professor, Institute of Policy and Planning Sciences

University of Tsukuba

1-1-1 Ten-nodai, Tsukuba, Ibaraki 305-8573, Japan

Phone: +81-298-53-5186

Fax: +81-298-55-3849

E-mail: tsutomu@sk.tsukuba.ac.jp

October 1999

ABSTRACT

The size-density hypothesis addressed in this paper is the relationship between facility size and facility density in time minimization theory. It is known that multi-Weber problems on a continuous plane show the relationship that facility density is proportional to demand density raised to the two-thirds power. Although it is rather hard to state this relationship in ordinary p -median problems on networks because of the difficulty in defining the notion of density, it is expected that the relationship holds also in these classical p -median problems. In this paper, we define the facility density at a certain node as the number of facilities which exist within a predetermined distance (radius) from that node, and test the hypothesis in p -median problems by calculating demand density and facility size, assuming a facility has the capacity to accommodate demand. Numerical tests will be examined to confirm this relationship.

Keywords: p -median problems, size-density hypothesis, location-allocation models

1. Introduction

The classical minisum location problems are to find the several locations that minimize the summed distance from discrete/continuous demands whose locations are known. The principal objective of previous studies on minisum location problems was to search an exact solution so far. For all practical purposes, however, it is not always necessary for us to know rigorous optimal locations, but often sufficient if we can only estimate the number of facilities for each area roughly. In that case, we only have to know the facility density that is necessary for the demand density at that place.

For instance, when we locate primary schools in a region, we usually allocate the number of schools according to population (or the number of pupils) and area of each subregion. Normally the subregion with high population density requires more schools than the average, and in a low population density subregion it is sufficient to have the number of schools below the average. If we locate schools in such a way that the summed distance from pupils to schools for all region is minimized, we first divide the targeted region into several subregions, and then allocate the number of schools to each subregion taking population into account. To what extent do we have to reflect population to the number of schools then?

Location problems may generally be classified into two categories: planar problems and network problems. In *planar problems*, minisum type location problems are formulated as *Weber* or *multi-Weber problems*. These problems are most often nonlinear optimization problems (e.g. Iri et al., 1984; Suzuki et al., 1991). In contrast, *network problems* are formulated as *median* or *p-median problems*. These problems tend to be linear zero-one optimization problems (see Daskin, 1995). Solution of these minisum problems generally becomes more difficult and time-consuming to find, as the number of facilities grows, especially for *p-median* problems. Therefore, it is rather hard to seek exact solutions.

Instead of relying on several heuristic methods, the *size-density hypothesis* might be useful in order to grasp the necessary facility density if we are not bound to exact optimal locations. The size-density hypothesis is the relationship between the facility sizes and the facility density in time minimization theory. It is known that multi-Weber problems on a continuous plane show the relationship that the facility density is proportional to the demand density raised to the two-thirds power. It is expected that the relationship holds also in these classical *p-median* problems. If it is true, we can estimate relatively easily the necessary number of facilities (or facility density) by using this relationship.

In this paper we present evidence that the size-density hypothesis almost holds in p -median problems and that it would be a useful idea to grasp the solution. Justification of the hypothesis is checked by calculating demand density and facility size for each node with a facility, assuming all facilities must have capacity to accommodate their demand that generates from the neighbourhood. Also, the relationship between demand density and facility size derived by the hypothesis is confirmed through some numerical examples.

It seems to be rather hard to state this relationship in ordinary p -median problems on networks because of the difficulty in defining the notion of density. In this paper we define the facility density at a certain node as the number of facilities that exist within a predetermined distance (radius) from that node. We try to test the hypothesis by calculating demand density and facility size assuming all facilities have the capacity to accommodate its demand.

The next section, Section 2, is a summary of the size-density hypothesis in multi-Weber problems. In Section 3, the exact solution of large p -median problems by linear programming relaxation is treated. Section 4 presents the test of the size-density hypothesis to p -median problems on a jurisdictional network of the Tokyo metropolitan area of Japan. Conclusion and further research agenda are presented in the last section.

2. Size-Density Hypothesis in multi-Weber Problems

The fact that the facility density should be higher for higher population density is generally recognized as being taken for granted. However, few researches have been seen which specify the relation of both quantitatively.

Palmer (1973) considered the situation to locate a number of service points in a district on a plane such that the mean travel distance is minimized, with the assumption that clients travel to their nearest points. He deductively showed that the density of the service points should be proportional to the two-thirds power of the population density.

Originally the size-density hypothesis was developed to account for variations in the size of territorial subdivisions such as counties, states, or regions. The terminology of the *size-density hypothesis* was introduced by Stephan (1972) who dealt with such a number of systems of administrative-territorial divisions including those in 98 modern countries. He examined the size of territorial units in relation to the population density of the units, which is derived under the theory of time minimization, and explained the empirical fact that countries usually locate their smaller

administrative subdivisions in regions of higher population density (Stephan, 1977). Stephan and McMullin (1981) concluded that the time-minimization theory also has been supported in the case of U.S. counties during the period of their formation.

Gusein-Zade (1982a, b) started with the discussion on the central place theory of Christaller and Lösch. He treated the Bunge's (1966) problem, which describes a deformation of the plane that leads to a constant population density (for more details, see Beavon, 1977; Rushton, 1972). Finally he concluded that the distribution of central places should be determined in such a way that the number of residents served by one center is proportional to the two-thirds power of the population density, instead of the one-half power Bunge had proposed. Stephan and Eggers (1985) have summarized the relation with some earlier works.

These findings are summarized as the theoretical relationships between facility density, facility size, area of zones, and demand density, all of which are based on the size-density hypothesis. The hypothesis is generally described as follows.

Let us consider that we place N facilities in a region D in which the population (demand) distributes with the density $h(\mathbf{x})$ at the location of $\mathbf{x} \in D$. We set the following assumptions as Stephan (1977) considered:

- (a) people receive services from the nearest facility,
- (b) the frequency of service use does not depend on the distance from the facility,
- (c) expenditures for travel distance are proportional to travel distance, and
- (d) expenditures for facility construction or operation do not depend on its size.

In the region D , the already given population density $h(\mathbf{x})$ is not supposed to be uniform generally. Also, facility density in the neighbourhood around $\mathbf{x} \in D$ (its area is given by dS) is denoted by $n(\mathbf{x})$. Let us assume that the number of facilities is sufficiently large and the facility density is uniform within a sufficiently small local area. Then the mean distance $d(\mathbf{x})$ to a facility around \mathbf{x} should be approximately given by

$$d(\mathbf{x}) = k/\sqrt{n(\mathbf{x})} \quad (1)$$

where k is a constant value. If the area served by one facility forms a regular hexagon, k

is given by $k = \frac{4 + 3 \log 3}{24} \left(\frac{2\sqrt{3}}{3} \right)^{3/2} \approx 0.3772$, and if the area forms a circle, k is given

by $k = \frac{2}{3\sqrt{\pi}} \approx 0.3761$. Suppose that we have total population H , which can be written

as

$$H = \int_D h(\mathbf{x})dS \quad (2)$$

Then the problem that minimizes the mean travel distance to the nearest facilities can be described as the following functional minimization problem:

$$\begin{aligned} \min_{n(\mathbf{x})} \quad \bar{d}[n(\mathbf{x})] &= \frac{1}{H} \int_D d(\mathbf{x})h(\mathbf{x})dS \\ &= \frac{k}{H} \sum_i \int_{D_i} \frac{h(\mathbf{x})}{\sqrt{n(\mathbf{x})}} dS, \\ \text{s.t.} \quad \int_D n(\mathbf{x})dS &= N \end{aligned} \quad (3)$$

where D_i is the territory of i -th facility and N is total number of facilities. The solution of this functional optimization can be obtained by using Lagrangean multiplier method. We then set Lagrangean as

$$L[n(\mathbf{x})] = \int_D \left\{ \frac{kh(\mathbf{x})}{H\sqrt{n(\mathbf{x})}} + \lambda n(\mathbf{x}) \right\} dS \quad (4)$$

where λ is the Lagrangean multiplier. From the calculus of variations, we can derive from the necessary condition that the following equation should be satisfied.

$$-\frac{k}{2H} \frac{h(\mathbf{x})}{\{n(\mathbf{x})\}^{2/3}} + \lambda = 0 \quad (5)$$

Rewriting the above equation, we obtain

$$n(\mathbf{x}) = C_1 \{h(\mathbf{x})\}^{2/3} \quad (6)$$

where C_1 represents a constant. This equation shows that, in optimum, *facility density* is proportional to population density raised to the two-thirds power, and that brings about minimum average travel distance. This fact means that the required number of facilities in a dense area is less than that in a sparse area even when population of the two areas is same.

Equation (6) is the most essential expression of the size-density hypothesis, but we can adopt other expressions. Since *the area served by one facility*, $a(\mathbf{x})$, is inversely proportional to facility density, therefore it should be proportional to population density

raised to the minus two-thirds power:

$$a(\mathbf{x}) = C_2 \{h(\mathbf{x})\}^{-2/3} \quad (7)$$

where C_2 is a constant. The size of population served by one facility (or *facility size*, if we assume facilities are not capacitated) is given by multiplying the area served by one facility by population density. Thus we have

$$s(\mathbf{x}) = a(\mathbf{x})h(\mathbf{x}) = C_3 \{h(\mathbf{x})\}^{1/3} \quad (8)$$

where $s(\mathbf{x})$ indicates facility size and C_3 is a constant.

In the real world, the assumptions (a)-(e) do not always hold. The size-density relationships derived above are not realized for such reasons that demand for facilities depends on the distance traveled, or that the travel cost is not proportional to the distance traveled. Some researchers have made many efforts for extensions, generalizations, and empirical analyses of the size-density hypothesis. For example, Stephan (1988) reexamined the exponent of travel cost and reformulated the hypothesis with the distributions of nongovernmental service centers. Gusein-Zade (1993) generalized the exponential functions with the exponent not equal to the number shown above.

3. Large p -Median Problems

For network location problems such as p -median problems, it may be prohibitively expensive to work with the necessary network data as well as its solution. In such cases, network location problems may well be approximated using planar location problems. Instead of distance on the network, we use Euclidean or rectilinear distance. The resulting problems are often easier to analyze.

On the other hand, many realistic assumptions can be incorporated in network location problems which cannot be included in planar location problems. In network models we have a transportation network which may represent a system of highways or railways. The distance between two points, which is usually defined as the shortest distance on a network, often represents actual spatial structure more accurately in network models than in planar models. Sometimes this fact tends to facilitate for us to treat p -median problems rather than multi-Weber problems, even when the problem is

quite large.

A great deal of endeavor should be necessary to do with large p -median problems. The methods that can be used for p -median problems are relaxed linear programming (ReVelle and Swain, 1970; Morris, 1978), the dual ascent methodology, Lagrangean relaxation (Galvão and Raggi, 1989), and so forth. However, there is a limit to these methodologies.

In that case, the size-density hypothesis is very useful to find the solution, if it holds in p -median problems. By looking at the demand density in the neighborhood of a certain node, we can estimate the number of facilities needed near that node. The rest of this paper is devoted to ascertain whether the hypothesis holds in p -median problems through some numerical examples.

ReVelle and Swain (1970) dealt with the method to solve p -median problems by relaxing the integer constraint to a simple non-negativity requirement. Morris (1978) also considered an application of linear programming (LP) to fixed charge plant location problems, which bear some resemblance to p -median problems. He found that ordinary linear programming typically produces integer solutions to uncapacitated problems. Rosing, ReVelle, and Vogelaar (1979) considered how to reduce both the computer space and computational time in the LP formulation of p -median problems. Rosing, Hillsman, and Vogelaar (1979) compared LP solution with other solving techniques including some heuristics.

In this paper the exact optimal solution, which is used for checking the hypothesis, is found by using the LP formulation of ReVelle and Swain (1970). The integer constraint on the decision variables in this formulation is relaxed to a nonnegativity constraint, and the presence of integers in the solutions is used as a test of feasibility.

4. Tests for the Size-Density Hypothesis in p -Median Problems

Equation (8) is one of the theoretical expressions of the size-density hypothesis. We use it to check whether the hypothesis holds also in p -median problems, by verifying the relationship between facility size and demand density through a numerical example.

Here we take up a large p -median problem with 280 demand points (simultaneously being candidates for facility locations), which represent local government offices in the Tokyo region of Japan. Let us assume that every node is

weighted by the 1990 population (assumed equivalent to demand) of their jurisdictions from the census (Figure 1). The area of circles in the figure indicates the population. The spatial linkages are given by Delaunay network or minimum spanning tree (MST) generated from the demand points, as shown in Figure 2 and 6, respectively. The reason why we focus our mind on these two networks is that the former represents the most similar situation to the Euclidean space in multi-Weber problems and the latter represents the situation of least similarity, which is farthest from multi-Weber problems. The locational configuration would depend on the network. We adopt the two typical cases. The distances are the inter-jurisdictional distances on the network measured by using latitude-longitude data for the locations of local government offices.

In the Delaunay network case (Figure 2), we have obtained (relaxed) LP solutions which fully terminated integer when $p=10, 19, 28, 40, 56,$ and 70 . Selected sites of a 28-median example are listed in Table 1, and illustrated by closed circles in Figure 3 in which assignments are expressed by minimum path trees. It is readily seen that higher facility densities take place in the vicinity of higher population densities (compare with Figure 1). Table 1 also indicates that a facility at higher population density area tends to have much demand and many demand points assigned.

To ascertain this relation quantitatively, let us define the notion of demand density on a network. As shown in Figure 4, for the demand point located at node j , consider a set N_j of nodes to which the shortest path distance from j does not exceed the given nonnegative covering radius, r . In that case, the demand density at node j , h_j , should be represented by

$$h_j = \frac{\sum_{i \in N_j} H_i}{\pi r^2} \quad (9)$$

where H_i is the population at node i . If we have dense network in all directions, h_j means the population density in the circle with radius r on a plane.

With this form of demand density and facility size s_j , let us test the following relation

$$s_j = C_3 h_j^\alpha \quad (10)$$

at each node j selected as a median, and check if the exponent α be approximately one-third by ordinary least squares (OLS) method (see Equation (8)). Figure 5 shows the

relationship between demand density and facility size in the 28-median example. The radius r was set to be 15 kilometers (approximately equal to the radius of average area per facility). Facility size does not increase so drastically as demand density increases. The regression result gives us the following relation:

$$s_j = 346168 h_j^{0.345} \quad (R^2=0.761) \quad (11)$$

The correlation shows a relatively high degree, and the estimated exponent $\hat{\alpha}$ is found very close to one-third as expected. It can be said that the size-density hypothesis almost holds in this case.

In the MST case (Figure 6), we again could obtain integer solutions by LP relaxation when $p=10, 19, 28, 40, 56,$ and 70 , and solutions illustrated by closed circles in Figure 7 (tree shows assignments). The tendency that higher facility density takes place in the vicinity of higher population density does not seem to change. The radius r is set to be 15 kilometers again, and the relationship between demand density and facility size in the 28-median solution is shown in Figure 8. As we can see, the relationship is not so good as that in the Delaunay case. The regression result provides us the following relation:

$$s_j = 342030 h_j^{0.421} \quad (R^2=0.646) \quad (12)$$

The estimated exponent $\hat{\alpha}$ becomes larger than in the Delaunay case. This is explained by the MST network being much less than two-dimensional space than the Delaunay network. This observation means that the difference with the degree of the connection of the network influences the robustness of the relation even if the same demand distribution is given.

In Tables 2 and 3, regression results with other numbers of facilities p are provided. Except for $p=10$ (in this case, the assumption of uniform demand distribution would not hold), we can find the fact that the estimated exponent $\hat{\alpha}$ is stable to be around one-third in the Delaunay cases. Also, even in the MST cases, we can confirm that $\hat{\alpha}$ is given steadily between 0.4 and 0.5, though the stability in the relation is not so good. The relationship is robust with regard to the number of facilities. This ensures that the size-density hypothesis holds true even in p -median problems.

5. Concluding Remarks

We have examined the size-density hypothesis in p -median problems. Seeking the exact solution obtained by applying LP relaxation to sample problems, it was shown that the relation that facility size is proportional to demand density raised to the one-third power also sufficiently holds in large p -median solutions, especially when the Delaunay network is given. This result may allow us to expect that we can solve large p -median problems in a relatively simple manner with some certainty for optimality.

This study is a start, by no means an end, to a further research agenda. First, the numerical example treated in the paper was only for typically ideal networks. It would be necessary to examine more realistic or non-typical networks. For future research, it might be necessary to reconsider the definition of demand density on networks. Second, the development of new heuristics for large p -median problems might be considered, though the method to seek the solution is not discussed in this paper. By grasping the distribution of the demand density in a targeted region and dividing into p areas in a way that each area has the demand scale as proportional to the one-third power, we can have a good approximation of the p -median solution. Third, we should verify whether the hypothesis could be proved empirically from actual facility locations (see Suzuki, 1999). Tests should be applied to various types of facilities, since many kinds of facilities would have different characteristics. This paper hopefully leads to these researches.

Acknowledgements

The earlier version of this paper was presented at the ISOLDE VIII, June 23-29, 1999, Coimbra and Estoril, Portugal. Thanks are due to Professors Pierre Hansen, Takeshi Koshizuka, Yoshiaki Ohsawa, and Shou Ohtsu for their valuable comments. This work was partially funded by the Grant in Aid for Scientific Research awarded by the Ministry of Education, Science, Sports and Culture of Japan under grant 10780273. This support is also gratefully acknowledged.

References

- Beavon, K.S.O. (1977) *Central Place Theory: A Reinterpretation*, Longman.
- Bunge, W. (1966) *Theoretical Geography*, Lund Studies in Geography, Series C,

Number 2, Lund: G.W. Gleerup.

- Daskin, M.S. (1995) *Network and Discrete Location: Models, Algorithms, and Applications*, Wiley.
- Galvão, R.D. and L.A. Raggi (1989) "A Method for Solving to Optimality Uncapacitated Location Problems," *Annals of Operations Research*, 18, 225-244.
- Gusein-Zade, S.M. (1982a) "Models of Region Size in a Hierarchical Administrative-Territorial System," *Soviet Geography*, 22, 599-607.
- Gusein-Zade, S.M. (1982b) "Bunge's Problem in Central Place Theory and its Generalizations," *Geographical Analysis*, 14, 3, 246-258.
- Gusein-Zade, S.M. (1993) "Alternative Explanations of the Dependence of the Density of Centers on the Density of Population," *Journal of Regional Science*, 33, 4, 547-558.
- Iri, M., K. Murota and T. Ohya (1984) "A Fast Voronoi-diagram Algorithm with Applications to Geographical Optimization Problems," in P. Thoft-Christensen (ed.) *Lecture Notes in Control and Information Science*, Vol. 59: *System Modelling and Optimization, Proceedings the IFIP Conference on System Modelling and Optimization*, Copenhagen, Springer, Berlin, 273-288.
- Morris, J.G. (1978) "On the Extent to Which Certain Fixed-Charge Depot Location Problems Can be Solved by LP," *Journal of Operational Research Society*, 29, 1, 71-76.
- Palmer, D.S. (1973) "The Placing of Service Points to Minimize Travel," *Operational Research Quarterly*, 24, 121-123.
- ReVelle, C.S. and R.W. Swain (1970) "Central Facilities Location," *Geographical Analysis*, 2, 30-42.
- Rosing, K.E., E.L. Hillsman and H. Vogelaar (1979) "A Note Comparing Optimal and

- Heuristic Solutions to the p -Median Problem,” *Geographical Analysis*, 11, 1, 86-89.
- Rosing, K.E., C.S. ReVelle and H. Vogelaar (1979) “The p -Median and its Linear Programming Relaxation: An Approach to Large Problems,” *Journal of Operational Research Society*, 30, 9, 815-823.
- Rushton, G. (1972) “Map Transformations of Point Patterns: Central Place Patterns in Areas of Variable Population Density,” *Regional Science Association Papers*, 28, 111-129.
- Stephan, G.E. (1972) “International Tests of the Size-Density Hypothesis,” *American Sociological Review*, 37, 365-368.
- Stephan, G.E. (1977) “Territorial Division: The Least-Time Constraint Behind the Formation of Subnational Boundaries,” *Science*, 196, 523-524.
- Stephan, G.E. (1988) “The Distribution of Service Establishments,” *Journal of Regional Science*, 28, 1, 29-40.
- Stephan, G.E. and M.L. Eggers (1985) “Bunge’s Problem in Central Place Theory and Its Generalizations: A Comment,” *Geographical Analysis*, 17, 3, 257-258.
- Stephan, G.E. and D.R. McMullin (1981) “The Historical Distribution of County Seats in the United States: A Review, Critique, and Test of Time-minimization Theory,” *American Sociological Review*, 46, 907-917.
- Suzuki, T. (1999) “Theoretical Relationship Between Facility Density and Population Density Determined by Minimizing Travel Cost of Users,” *Journal of Architecture, Planning and Environmental Engineering (Transactions of Architectural Institute of Japan)*, 521, 183-187. [in Japanese]
- Suzuki, T., Y. Asami and A. Okabe (1991) “Sequential Location-Allocation of Public Facilities in One- and Two-Dimensional Space: Comparison of Several Policies,” *Mathematical Programming Series B*, 52, 1, 125-146.

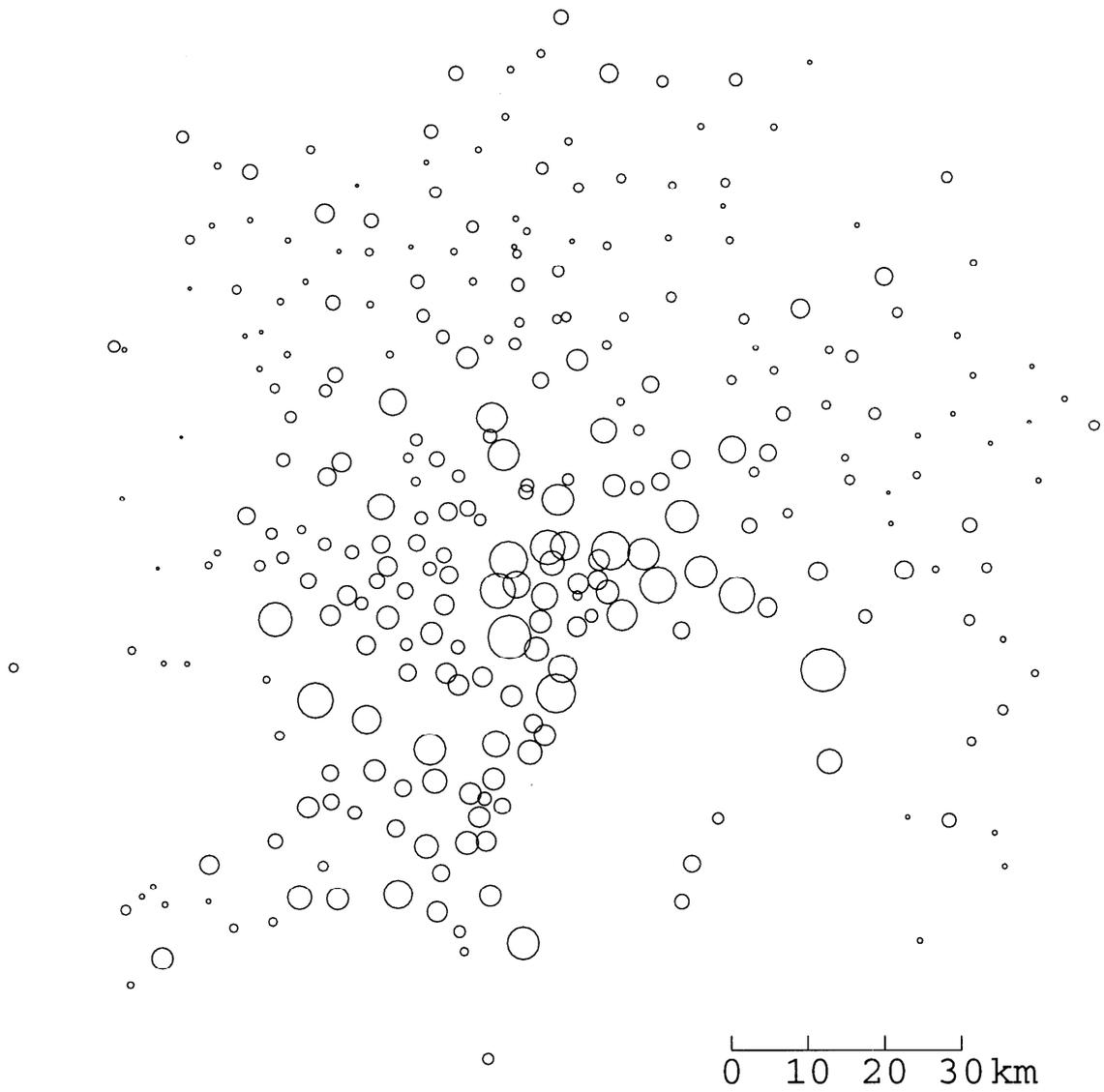


Figure 1. Population distribution in 1990 in the Tokyo region.

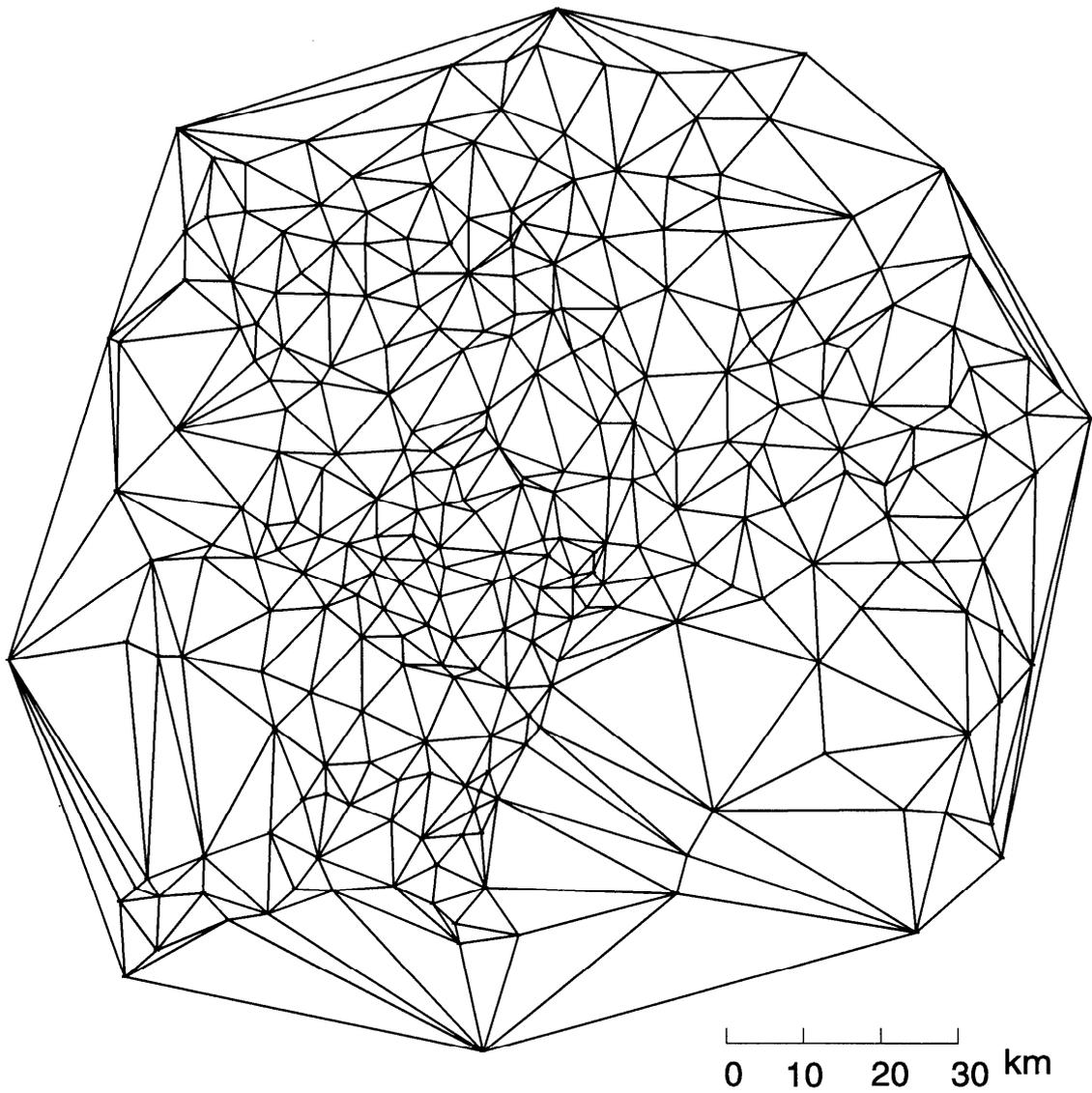


Figure 2. Delaunay network generated by the location of local government offices in the Tokyo region.

Table 1. A p -median solution on the Delaunay network in the Tokyo region ($p=28$).

No.	Jurisdiction name	# of assigned nodes	Demand size	Population density $r=15$
1	Tsuchiura	16	624605	6.23
41	Oyama	17	761817	5.73
49	Kawagoe	14	1105247	20.78
50	Kumagaya	28	1015496	8.09
52	Urawa	7	1224982	42.76
56	Tokorozawa	12	1251780	39.51
70	Koshigaya	6	732005	28.31
120	Miyashiro	22	953377	14.49
129	Chiba	8	1301222	20.55
131	Funabashi	6	1474072	41.74
132	Kisarazu	4	278256	3.75
141	Kashiwa	13	1024252	23.38
152	Shisui	17	585865	6.25
177	Sumida	8	1396987	112.44
181	Ota	5	1519840	70.32
182	Setagaya	6	1646904	101.31
184	Nakano	5	1800088	114.15
189	Itabashi	4	1558600	116.22
192	Katsushika	5	2139768	94.06
194	Hachioji	16	1154351	31.84
203	Koganei	14	1771868	64.65
227	Yokohama-Kanagawa	7	1334701	59.88
238	Yokohama-Midori	6	1288119	59.82
240	Yokohama-Sakae	8	1599623	50.21
249	Yokosuka	4	571727	16.72
253	Odawara	9	490371	6.94
256	Sagamihara	5	1047205	35.07
267	Samukawa	8	993474	25.11
Total		280	32646602	

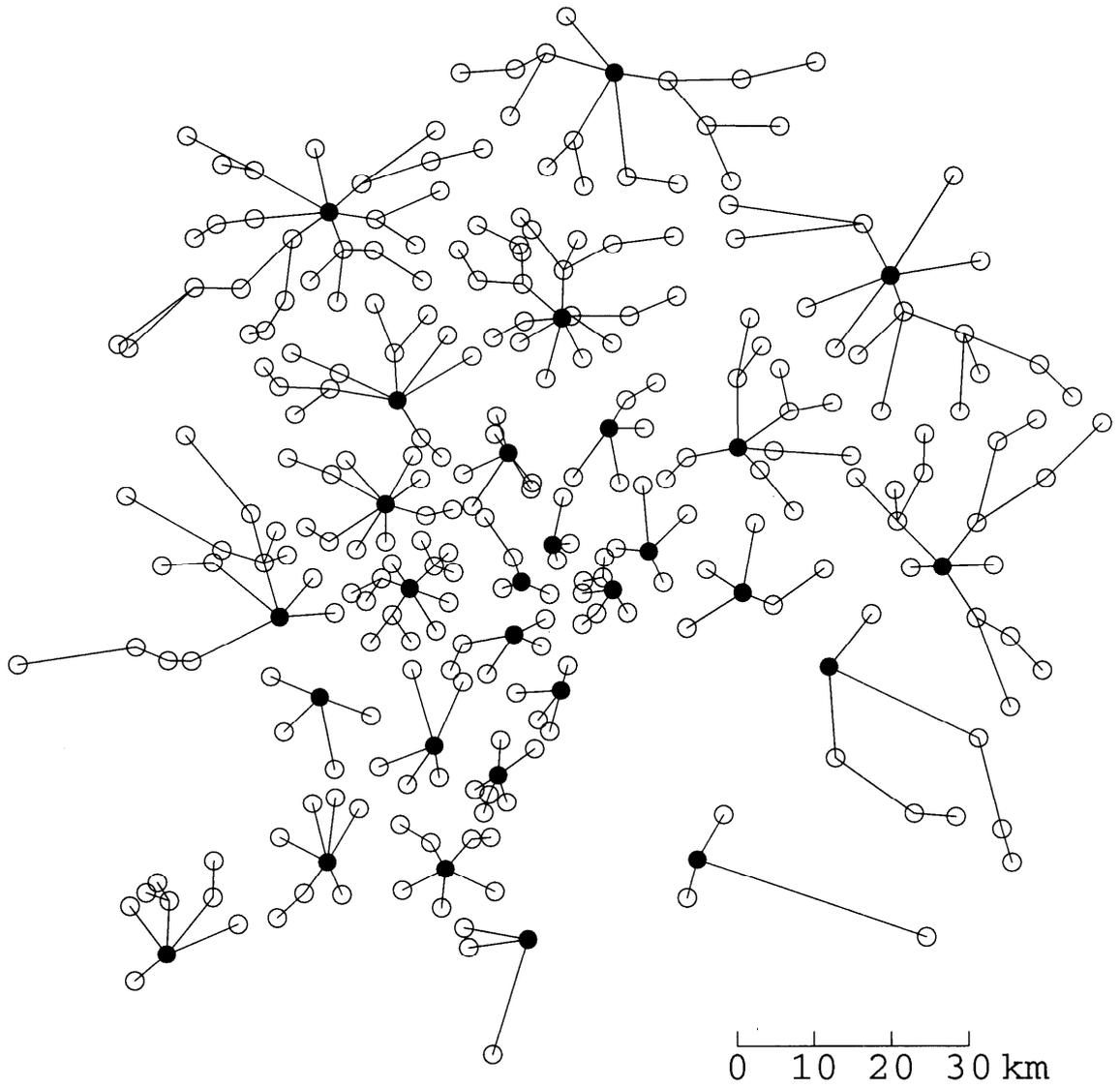


Figure 3. A p -median solution on the Delaunay network ($p=28$).

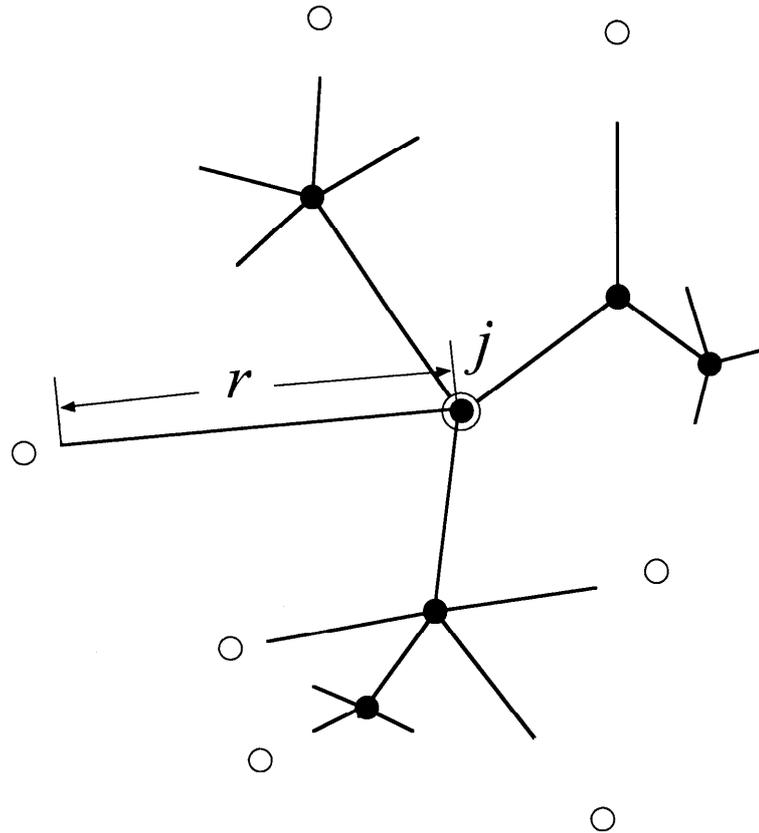


Figure 4. Definition of demand density on a network.

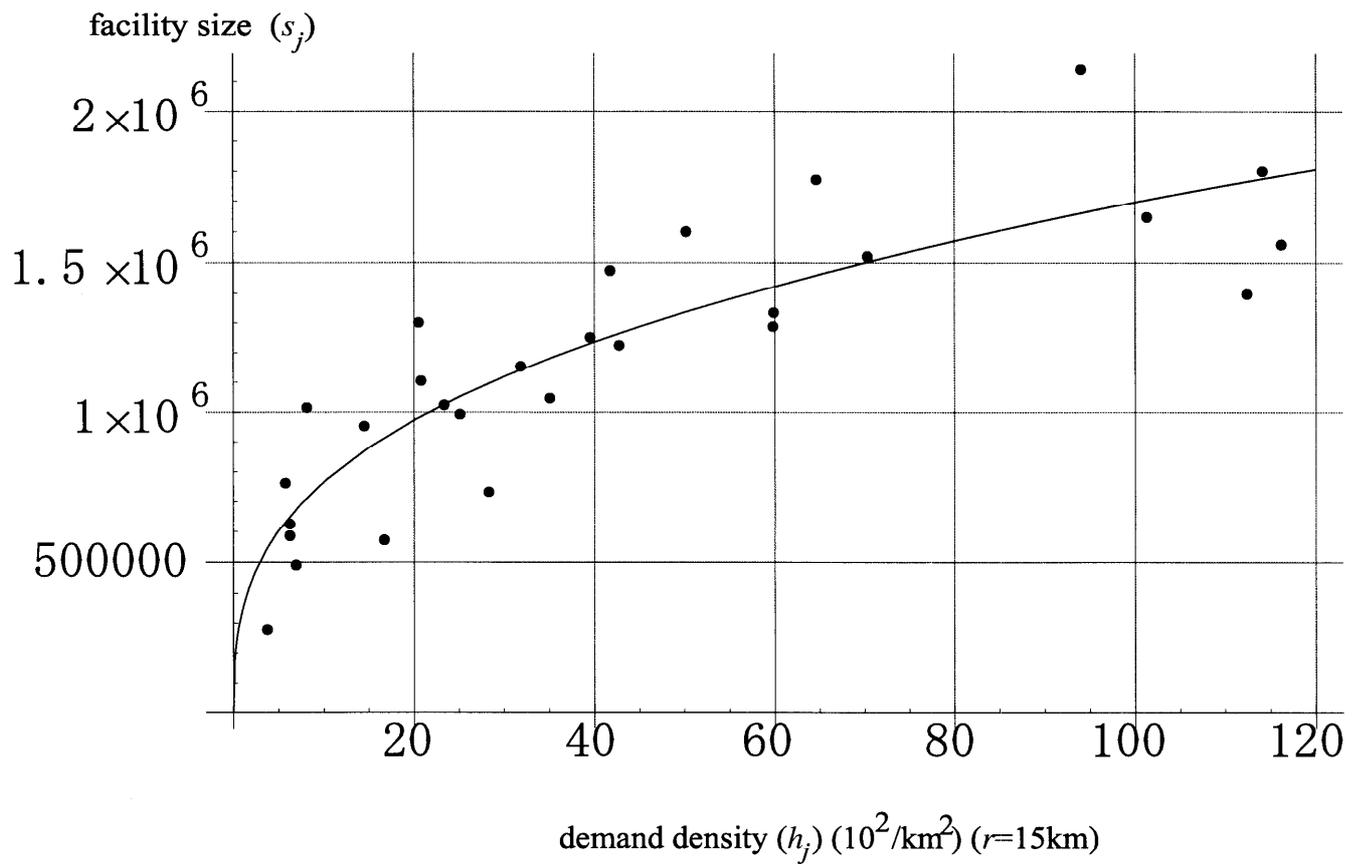


Figure 5. Size-density relationship in the solution on the Delaunay network ($p=28$).

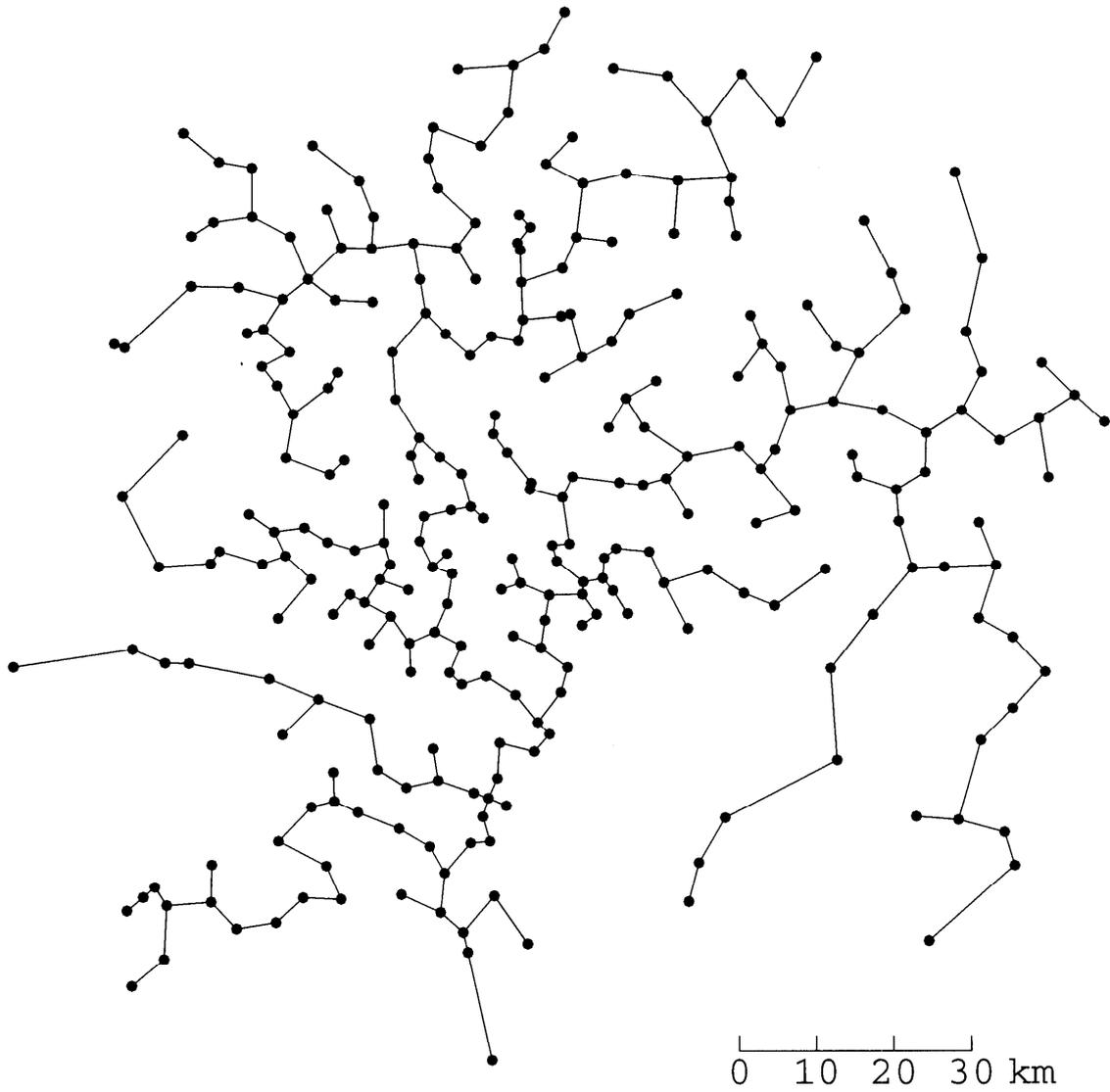


Figure 6. Minimum spanning tree (MST) generated by the location of local government offices.

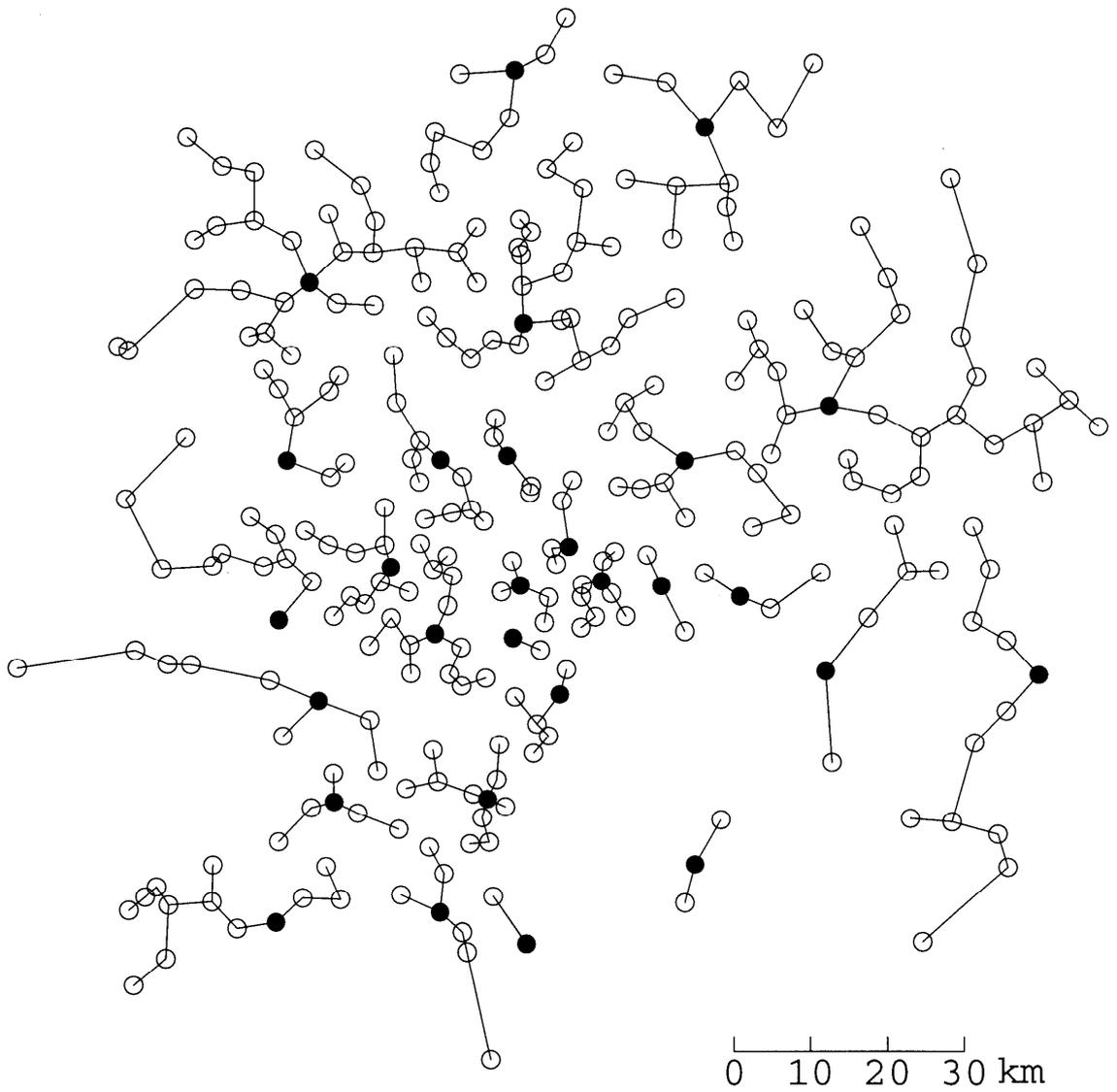


Figure 7. A p -median solution on the MST network ($p=28$).

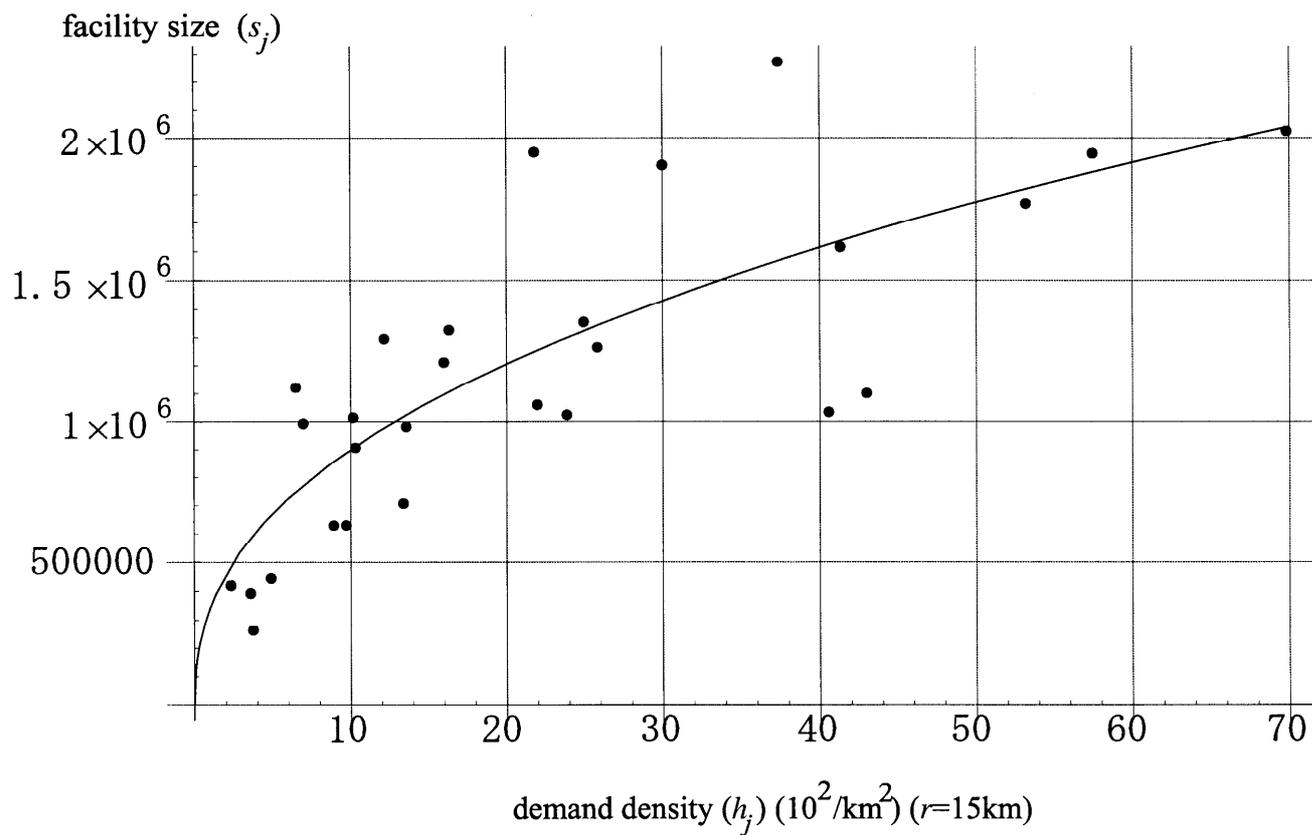


Figure 8. Size-density relationship in the solution on the MST network ($\varphi=28$).

Table 2. Regression results of the size-density relationship on the Delaunay network.

p	r (km)	\hat{C}_3	$\hat{\alpha}$	R^2
10	20	338951	0.642	0.94489
19	15	447609	0.375	0.74489
28	15	346168	0.345	0.76098
40	10	249245	0.329	0.66853
56	10	160859	0.365	0.60495
70	10	138109	0.344	0.52008

Table 3. Regression results of the size-density relationship on the MST network.

p	r (km)	\hat{C}_3	$\hat{\alpha}$	R^2
10	20	865514	0.451	0.58461
19	15	380288	0.496	0.75266
28	15	342030	0.421	0.64632
40	10	159401	0.507	0.74844
56	10	134673	0.474	0.71685
70	10	146284	0.380	0.55419