Department of Policy and Planning Sciences

Discussion Paper Series

No.1363

Spatial Analysis on Accuracy of Travelling Distance on Network

by

ZHONG Dai, TAMURA Kazuki, OHSAWA Yoshiaki

November 2019

UNIVERSITY OF TSUKUBA
Tsukuba, Ibaraki 305-8573
JAPAN

Spatial Analysis on Accuracy of Travelling Distance on Network

ZHONG Dai
Institute of Policy and Planning Sciences, University of Tsukuba
TAMURA Kazuki
Asian Growth Research Institute
OHSAWA Yoshiaki
Institute of Policy and Planning Sciences, University of Tsukuba

Abstract

It is usual that Traffic analysis deals with data aggregated at representative points instead of dealing the whole existing data. As a reason, there is an urgent need of expenditure and high technology to dealing with detailed data. Another reason is that the estimate of traffic of logistics or human is always linked with other socioeconomic analysis and the data, of which could be more uncontrollable and changeable. Furthermore, in recent years, from the viewpoint of protection of personal information, the precision of the collected data may be dropped and cause the aggregation of the collected data. Therefore, in order to propose a way to grasp the error between the two kinds of data, this work will build a model which represent the two kinds of data with the continuous distance and discrete distance between two points in the network to simulate the data before aggregated and the aggregated data, and further discuss the gap of them, aiming at deriving the upper and lower limits of the gap.

Keyword：Continuous Distance，Discrete Distance，Traffic Data，Road Network，Crofton's formula

Contents

1.Introduction

As expounded in the Abstract, there are many reasons which could cause gap between original data and analytical data. And they could be roughly divided in two: the data could not be collected for some inevitable factors such as time or economic cost; and the other one is that a part of data would be discarded from collected data in order to keeping the data quality, reducing the cost of processing or protecting the privacy for data acquisition objects. Both of those reasons will have caused the aggregation of data. And this work will build a model to study in the gap caused by aggregation.

2.Tow kinds of distance in the network

2-1 Average distance between two points

In this work, two kinds of methods are defined to measure the distance between two points. The first one is to measure the average distance of two points which is randomly distributed in the network. Dividing the network in equal segments, and using midpoints to represent segments, the other one is to measure the average distance between those midpoints which represent those segments. And in all the figures of this work, all of the randomly distributed points are displayed in white points while those midpoints are displayed as black points (fig.1).

Fig.1 continuous points and discrete points

As expounded in the introduction, in order to look into the error between the data before aggregated and the aggregated data, this work transform the error into a calculable mathematical problem. And research on this error could be able to be discussed by the analysis of the gap between the two kinds of methods to measure the distance between two points in the network.

2-2 Antipodal and antipodal points

Before the introduction of the calculation, there is another conception defined in this work should be clarified. Between two point distributed in the network, there are always two pathways between them. If the two pathways have the same length, this pair of points are defined as antipodal points. And if the segments in the network are completely consisted of antipodal points, those pair of segments are defined as antipodal zones. Further take consider of double count, this work defined λ as double of the amount of antipodal zones pairs.

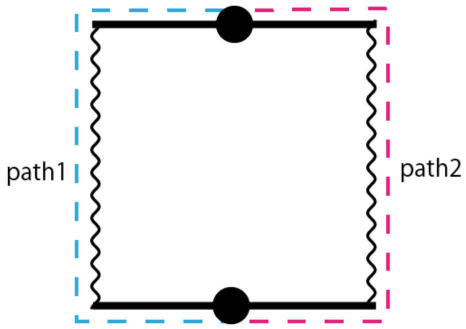


Fig.2　antipodal points: path1=path2

For example, in the 1x2 lattice-network, there are 5 pairs of antipodal zones while λ will be 10 in this situation.

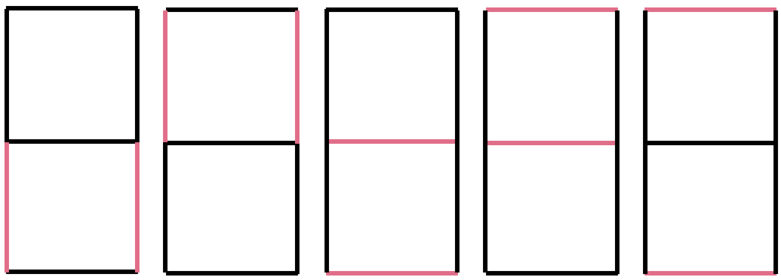


Fig.3 1x2 lattice-network

Furthermore, as the premise of the simulation model, all of the segments have the same length, and the pathways between two points are uniformly and independent in the network. Finally, all the antipodal zones are all contained by segments in networks.

3.The gap of the two kinds of distance

In this chapter, we will discuss about the calculation of the discrete average distance and continuous average distance. Then based on that we will derive the general formula of the gap of the two kinds of distance. Furthermore, in order to check the correctness of the formula, a series of inspection will be conducted later.

3-1 the discrete average distance and continuous average distance

In this work, it is assumed that the trip occurs in networks are uniformly and independently. In the situation of the discrete average distance, each segment is considered as one segment the point which represent that segment is placed at the middle of this segment. Let l_{ij} be the shortest distance between segment i and segment j, the amount of the segments be k, the average distance between those midpoints $\bar{d}$ (discrete average distance) will be presented as followed:

$$\bar{d} = \frac{1}{k^2} \sum_{i=1}^{k} \sum_{j \neq i}^{k} l_{ij}$$

When it comes to the continuous average distance, the method of calculation will be derived based on the result of discrete average distance. The average distance could always be divided as two parts in whatever discrete average distance or continuous average distance: the average distance between the pair of points which distributed in the same segment in the network (inner-distance). And the average distance between the pair of points which come from different segments in the network (approximately distance).

Firstly, in the situation of the calculation of the average inner-distance, for discrete average distance the inner distance will be zero while the continuous average distance is obviously not equal to zero (fig.4).

Fig.4 inner-distance

Supposed that there are two points(x_1,x_2) which are randomly distributed in the same segment which have a length of a. and the shortest distance between them is d (fig.5). And this distance could be represent in the coordinate system like fig.6. Causing that the average distance of all the random-distribution pairs of points will be $\frac{1}{3}a$ according to the fig.7. Hence, compared to the discrete average distance which the inner-distance equals to 0, the continuous average distance will be longer by $\frac{1}{3}a$ in every segment.

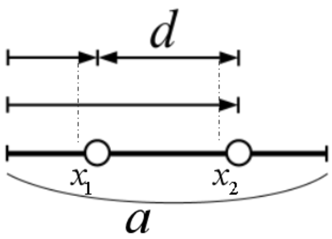
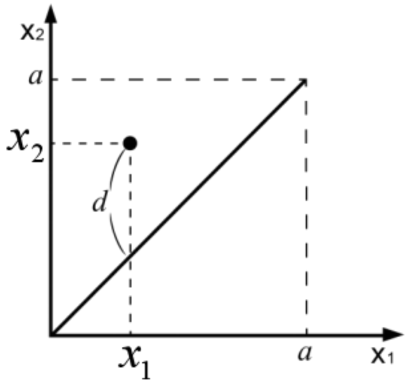


Fig.5 arbitrarily distribute two points



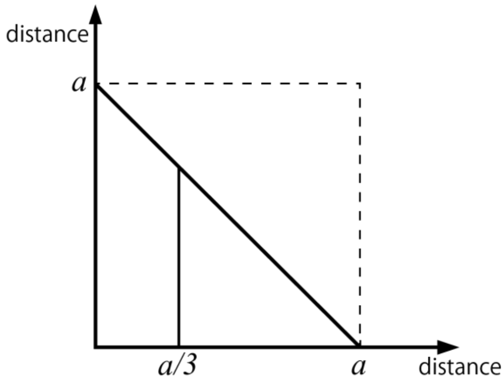


Fig.6 distance in coordinate system Fig.7 continuous average distance

And for the calculation of the approximately distance, the pathway between two points in networks are always be two, causing the shortest pathway between antipodal points is always shorter than that of the distance between two midpoints (fig.8).

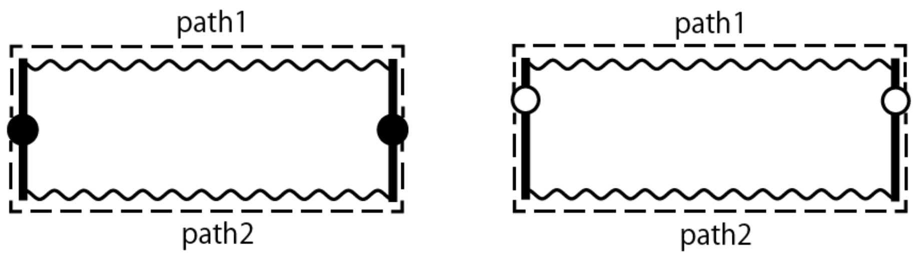


Fig.8 antipodal points and aantipodal points

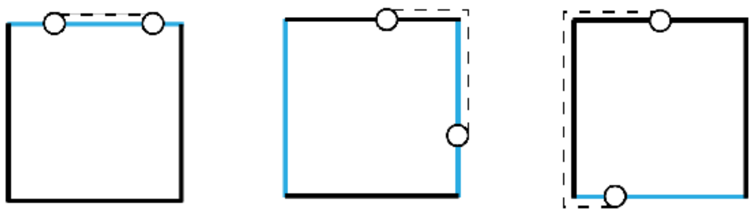
When calculate the approximately distance of continuous average distance based on that of discrete average distance, take the base unit in lattice-network for example. In the base unit, discrete average distance will be $2a$ while the continuous average distance in the whole unit will be a (fig.9).

Fig.9 discrete and continuous average distances in base unit

Taking apart the situation of the continuous average distance, the continuous average distance could be divided into 3 parts A, B and C (fig.10). And based on that the continuous average distance in part.A and part.B are $\frac{1}{3}a$ and a (referring to former calculation), and the occurrence accuracy of part.A, part.B and part.C are $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$. Let x be the continuous average distance of part.C, the equation could be established as followed:

$$a = \frac{1}{4} \times \frac{1}{3}a + \frac{1}{2} \times a + \frac{1}{4} \times x$$

It is not complicated that the continuous average distance of part.C equals to $\frac{5}{3}a$, which means that the continuous average distance between the pair of points which exist in antipodal segments is $\frac{5}{3}a$. Hence, compare to the discrete average distance which the approximately distance equals to $2a$, the continuous average distance will be shorter by $\frac{1}{3}a$ in every antipodal links pair.



Part.A Part.B Part.C

Fig.10 continuous average distance divided in 3 parts

Based on the result of incremental portion and reduce portion which has been calculated, the continuous average distance $\bar{r}$ could be presented as followed:

$$\bar{r} = \bar{d} + \frac{1}{3k}a - \frac{\lambda}{3k^2}a$$

Furthermore, the general formula of deviation could be stated as followed:

$$G \equiv \bar{r} - \bar{d} = \frac{k - \lambda}{3k^2}a$$

3-2 Check of the deviation formula

3-2-1 tree

Take the tree-network which has three segments A, B, C with the length of a for example.

In fig.11(a), midpoints represent the segments, and the distance between

them are represented in Table.1(a). From this table, we can calculate the average of them is $\frac{6a}{3^2}=\frac{2}{3}a$. In the situation of the calculation of the continuous average distance in this tree, the distance of the trips happen in every segments are represented as Table.1(b), and causing the average equals to $\frac{1}{3^2}\left(\frac{1}{3}a\times 3+6a\right)=\frac{7}{9}a$. By the way this work mark the increased parts of continuous average distance compared with discrete average distance as red while the reduced parts of continuous average distance compared with discrete average distance as blue in the calculation table.
The gap between them came out to be $\frac{1}{9}a$, which accordant with that calculated by general deviation formula: $G=\frac{k-\lambda}{3k^2}a=\frac{3-0}{3\times 3^2}a=\frac{1}{9}a$ because the number of the pairs of antipodal zones is 0.

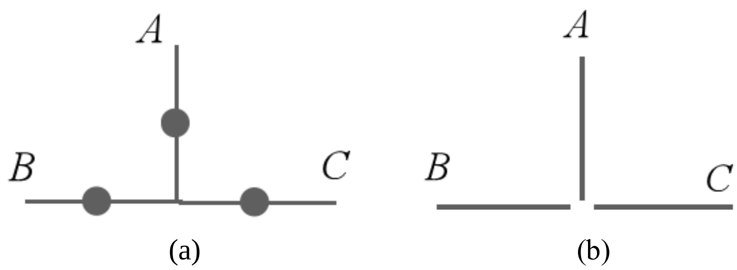


Fig.11 tree-network with 3 segments

	A	B	C
A	0	a	a
B	a	0	a
C	a	a	0

(a)

	A	B	C
A	$\frac{1}{3}a$	a	a
B	a	$\frac{1}{3}a$	a
C	a	a	$\frac{1}{3}a$

(b)

Table.1 discrete and continuous average distances in tree-network with 3 segments

3-2-2 circle

By the same token, in the circle which is divided as two equal parts. While $k=2$ and $\lambda=2$ the gap calculated by general deviation formula is $G=\frac{k-\lambda}{3k^2}a=\frac{2-2}{3\times 2^2}a=0$

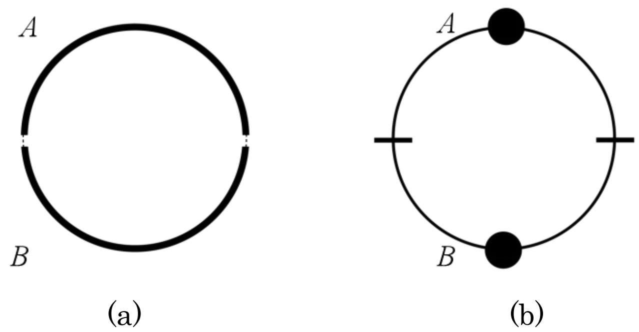


Fig.12 circle divided in two segments

	A	B
A	$\frac{1}{3}a$	$\frac{4}{3}a$
B	$\frac{4}{3}a$	$\frac{1}{3}a$

(a)

	A	B
A	0	a
B	a	0

(b)

Table.2 continuous and discrete average distances in circle

And according to Table.2, the gap between two kinds of average distance comes to be 0, witch is accordant to the result from general deviation formula.

3-2-3 lattice-network

In the situation of 1×2 lattice-network of which every segment have a length of a.(fig.13)

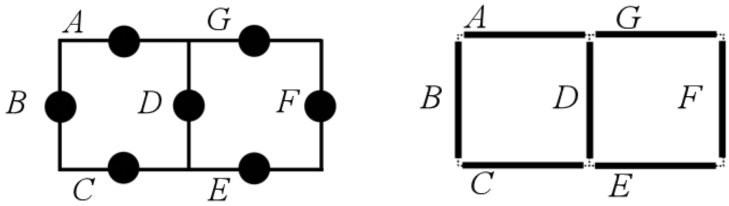


Fig.13 1x2 lattice-network

According to the calculation in the Table.3, the gap between them is $G = \bar{r} - \bar{d} = \frac{65}{49}a - \frac{66}{49}a = -\frac{1}{49}a$. In the 1×2 lattice network, $k = 7$, $\lambda = 10$, and the result of the calculation through the general deviation formula comes out to be $G = \frac{k-\lambda}{3k^2}a = \frac{7-10}{3\times 7^2}a = -\frac{1}{49}a$.

	A	B	C	D	E	F	G
A	$\frac{1}{3}$	1	$\frac{5}{3}$	1	2	2	1
B	1	$\frac{1}{3}$	1	$\frac{5}{3}$	2	$\frac{8}{3}$	2
C	$\frac{5}{3}$	1	$\frac{1}{3}$	1	1	2	2
D	1	$\frac{5}{3}$	1	$\frac{1}{3}$	1	$\frac{5}{3}$	1
E	2	2	1	1	$\frac{1}{3}$	1	$\frac{5}{3}$
F	2	$\frac{8}{3}$	2	$\frac{5}{3}$	1	$\frac{1}{3}$	1
G	1	2	2	1	$\frac{5}{3}$	1	$\frac{1}{3}$

(a)

	A	B	C	D	E	F	G
A	0	1	2	1	2	2	1
B	1	0	1	2	2	3	2
C	2	1	0	1	1	2	2
D	1	2	1	0	1	2	1
E	2	2	1	1	0	1	2
F	2	3	2	2	1	0	1
G	1	2	2	1	2	1	0

(b)

Table.3 discrete and continuous average distances in 1x2 lattice-network

3-2-4 general lattice-network

Furthermore, take the general lattice-network raised by Koshizuka[2] for consideration(fig.14).

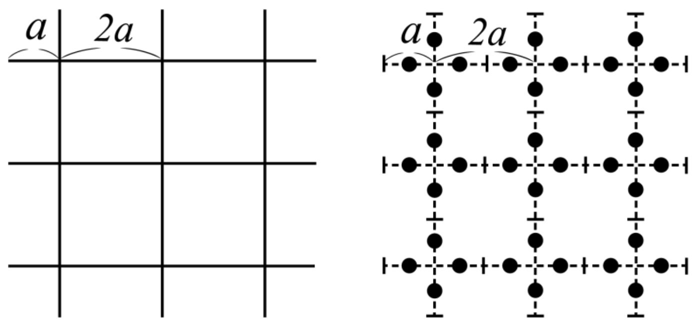


Fig.14 general lattice-network

According to the function graph of the distribution of the average distance on the network, continuous average distance come to be $\bar{r} = \frac{\int_0^{10b} rf(r)dr}{(36a)^2} = \frac{631}{162}a$. ($f(r)$ represent the function of the distribution of the continuous average distance on networks(fig.15).

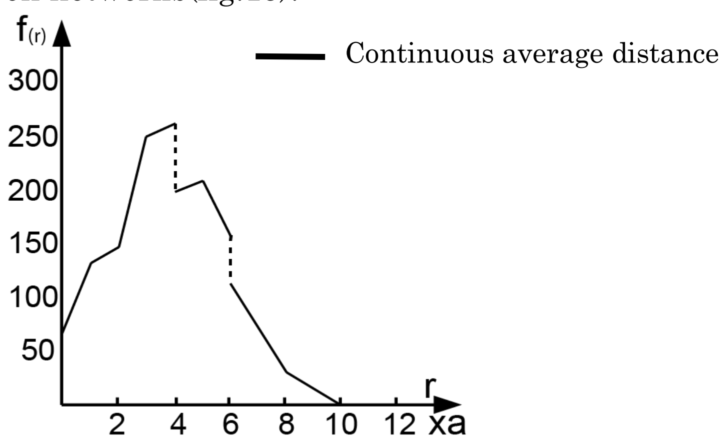


Fig.15 continuous average distance derived by Koshizuka

According to the symmetry of the lattice-network, discrete average distance could be figure out and $\bar{d} = \frac{1263}{324}a$, causing the gap of them $G = -\frac{1}{324}a$. In this lattice-network, $k = 36$, $\lambda = 48$ and the result of the calculation by the general deviation formula $G = \frac{k-\lambda}{3k^2}a = \frac{36-48}{3\times36^2}a = -\frac{1}{324}a$.

3-2-5 hexagon-network

In the hexagon-network (fig.16) the inspection will be conducted similarly.

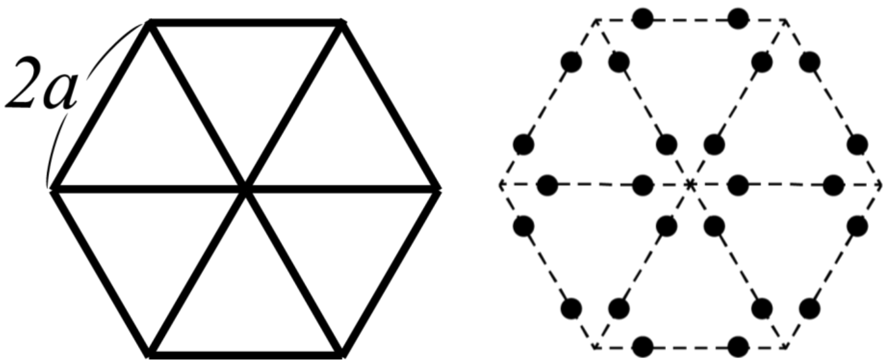


Fig.16 hexagon-network

In this situation, let's confirm theλfirstly. In the surrounding segments of the hexagon-network, there are six pairs of antipodal zones in each surrounding segments, $a - a_1$, $a - a_2$, $a - a_3$ and $b - b_1$, $b - b_2$, $b - b_3$. And in the inner segements, there are six pairs of antipodal zones in each inner segments, $c - c_1$, $c - c_2$, $c - c_3$ and $d - d_1$, $d - d_2$, $d - d_3$ (fig.17). and according to that, by the general deviation formula, the gap between the two kinds of distance is $G = \frac{k-\lambda}{3k^2}a = \frac{24-72}{3\times24^2}a = -\frac{1}{36}a$.

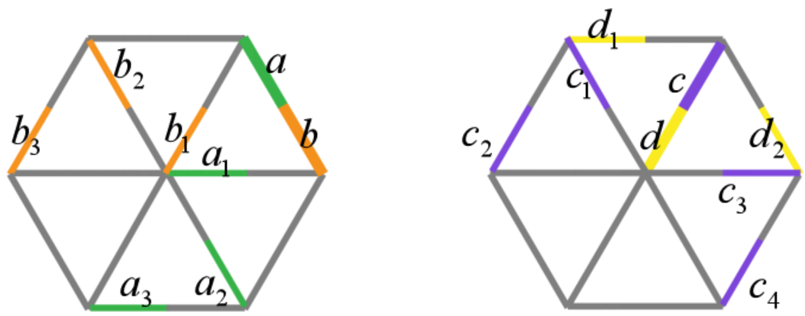


Fig.17 antipodal zones in hexagon-network

When calculate the discrete average distance for inspection, $\bar{d} = \frac{187}{72}a$ could be derived by symmetry. And in the situation of continuous average distance, similarly take the base unit as explanation for that.

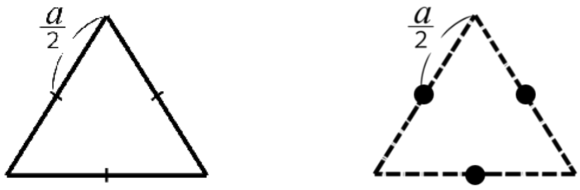


Fig.18 base unit of hexagon-network

In the triangular base unit of the hexagon-network(fig.18), the continuous average distance is a in the whole segments. And based on the result of the continuous average distance in the square, the continuous average distance in a single triangular unit $\bar{r}_1$ could be calculated as followed:

$$\bar{r}_1 = \frac{1}{4}\times\frac{1}{2}a + \frac{1}{2}\times\left(\frac{a}{2}+\frac{a}{2}\right) + \frac{1}{4}\times\left(a+\frac{2}{3}\times\frac{a}{2}\right) = \frac{23}{24}a$$

$\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$ are represent the probability of occurrence of the parts showed in fig.19

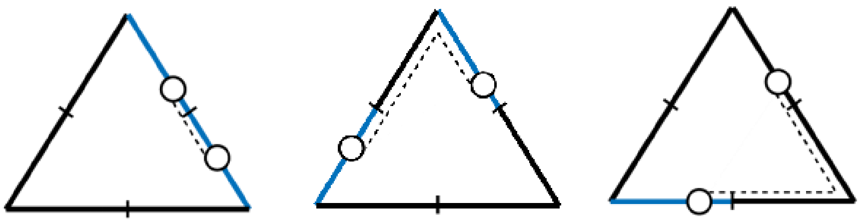


Fig.19 continuous average distance divided in 3 parts

Based on that, $\bar{r}$ in a hexagon-network could be calculated as $\frac{187}{72}a$. and the gap between them will be $G = \bar{r} - \bar{d} = -\frac{1}{36}a$, which is exactly the same with the result from the general deviation formula.

According to the check of the general deviation formula above, we could reach to the conclusion that the general deviation formula could apply for the tree, lattice, hexagon networks which have the same length as have been analyzed in examples.

4.The application of the general deviation formula

4-1 tree

In the shape of tree-networks, there will not have any pairs of antipodal zones, which means that the general deviation formula for tree-networks will be $G = \frac{1}{3k}a$.

From the general deviation formula, it can be tell that the deviation in tree-networks are only related with the numbers of segments. And the deviation is unrelated to the shapes of the tree-networks as long as they have the same number of segments(fig.20). And the deviation in tree-networks go to be underestimation.

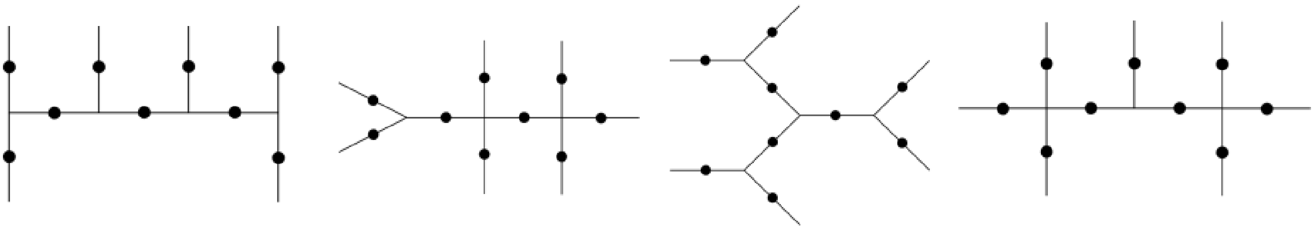


Fig.20 tree-networks, $k = 9$

4-2 lattice-network

In the general lattice-network, let the scale of the network be $m \times n$ (fig.21), the pairs of antipodal zones:

$$\lambda = C_{n+1}^{2} \times m \times 2 + C_{m+1}^{2} \times n \times 2 = mn(m+n+2),$$

And the number of segments:

$$k = (m+1)n + (n+1)m = 2mn + m + n$$

The general deviation formula for $m \times n$ lattice-network will be

$$G = \frac{k-\lambda}{3k^2} a = \frac{(m+n)(1-mn)}{3(2mn+m+n)^2} a$$

From the formula, it could be concluded that the deviation of lattice-networks go to be overestimation.

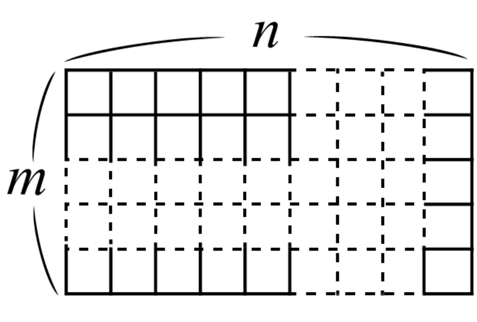


Fig.21 $m \times n$ lattice-network

4-3 the upper and lower limits of the general deviation formula

From the general deviation formula, the upper and lower limits could be derived out as:

$$-\frac{k-2}{3k} \leq G \leq \frac{1}{3k}a$$

And when remove the k from the general deviation formula, the upper and lower limits go to be as followed:

$$-\frac{1}{3}a < G \leq \frac{1}{3}a$$

In that, the upper limit will be acquired as the numbers of the segments are 1 and the pairs of antipodal zones are 0, which showed up as a 1-segment tree, referring to fig.22(a). And in the other hand, the lower limit will be acquired as there are infinity segments between two points, referring to fig.22(b).

(a) (b)

Fig.22 upper and lower limits

4-4 Subdivision in the network

When subdividing the segments in networks(fig.23), If s represent the number of the divided segments, the antipodal links pairs in the network will turn to s times compared with the number of antipodal links pairs when $s = 1$. So the general deviation formula will change to be as followed:

$$G_s = \frac{1}{s^2}G$$

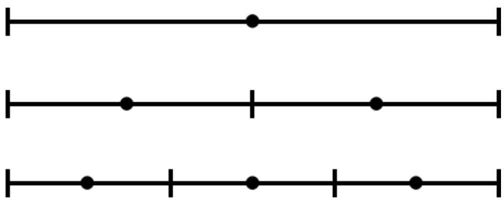


Fig.23 s=1, s=2, s=3

From this formula, the conclusion could be draw that the deviation between the discrete average distance and continuous average distance could decrease as the segments in the networks being subdivided.

So with the help of ArcMap10.6, the distance could be derived under the network analysis tool box to confirm the calculations. So the simulation results will be showed in next.

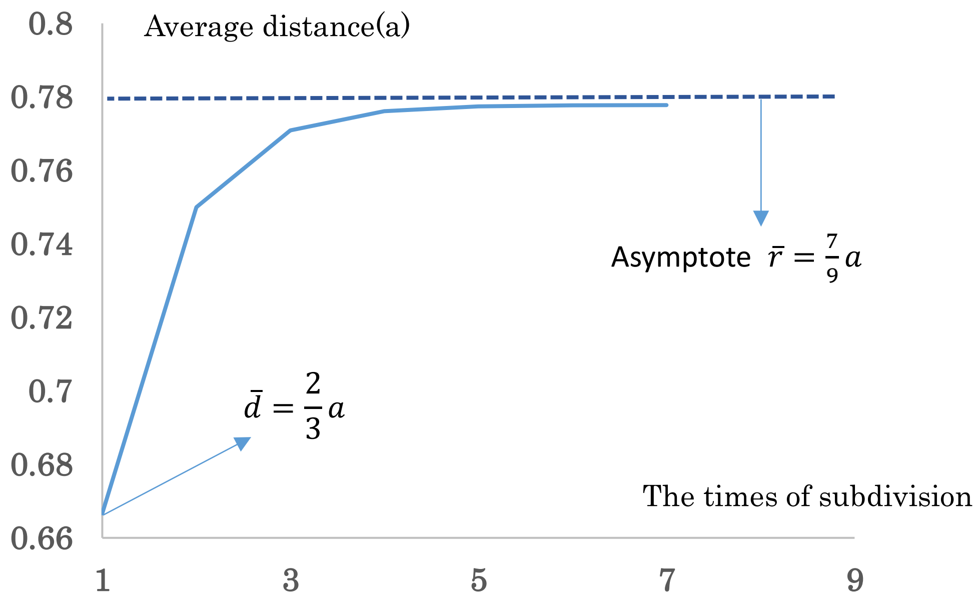


Fig.24 the average distance in tree(k=3), $\bar{d} = \frac{2}{3}a$, $\bar{r} = \frac{7}{9}a$

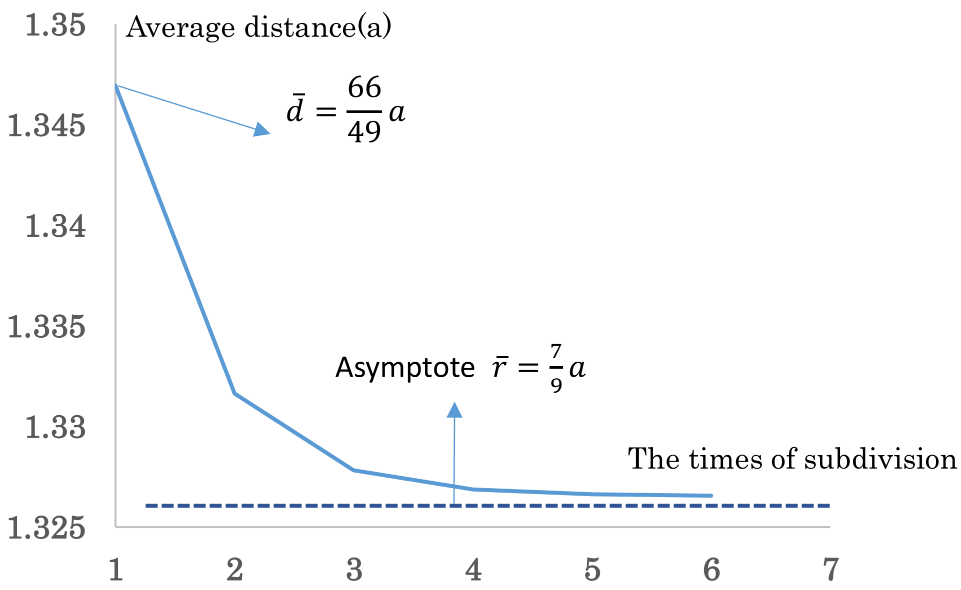


Fig.25 the average distance in lattice-networks(1x2), $\bar{d} = \frac{66}{49}a$, $\bar{r} = \frac{65}{49}a$

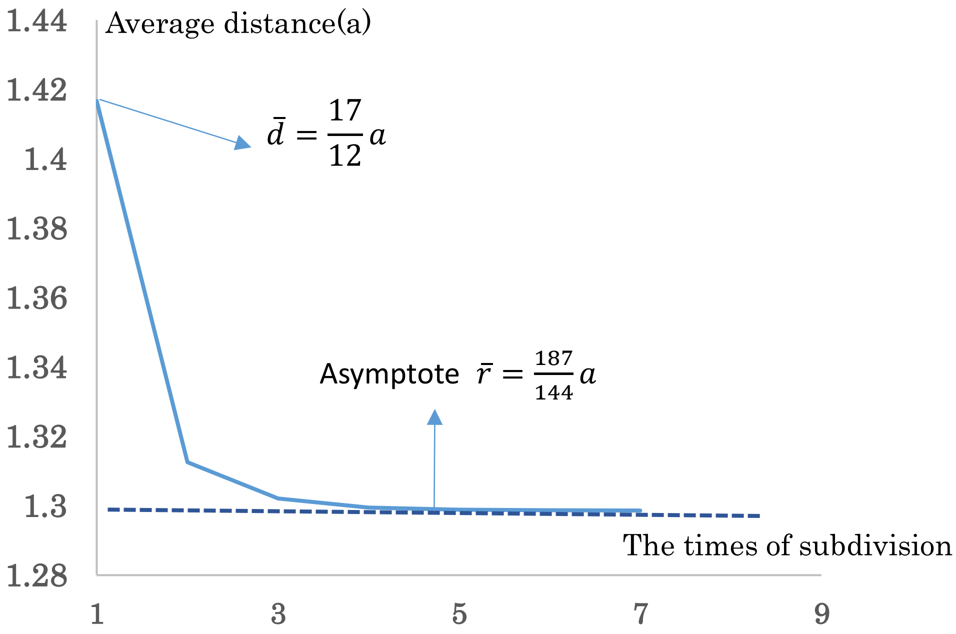


Fig.26 the average distance in hexagon-networks, $\bar{d} = \frac{17}{12}a$, $\bar{r} = \frac{187}{144}a$

5.The derive of gap by Crofton's formula

In the chapter.3, the general deviation formula has been derived as followed:

$$G \equiv \bar{r} - \bar{d} = \frac{k-\lambda}{3k^2}a \tag{1}$$

And in this chapter, this work will derive the general deviation formula by the using of Crofton's formula.

Increments will be joint in every segment in networks. Let the length of increments as $\varDelta$ in each segment, and A represent the original parts as B represent the increments(referring to fig.27). $\mu_{AA}, \mu_{AB}, \mu_{BB}$ respectively represent the inner-distance in part.A, the approximately distance between part.A and part.B, the inner-distance in part.B. Based on that, let $\mu(x)$ represent the inner-distance in segment which have the length of x. For $\bar{r} = \mu(a)$:

$$\mu(a + \varDelta) = \frac{1}{(a + \varDelta)^2}(a^2\mu_{AA} + 2a\varDelta\mu_{AB} + \varDelta^2\mu_{BB})$$

could be derived.

Furthermore, if the cross-section traffic volume from part.A to part.B is b, the inner-distance in part.A could be written as $\mu_{AA} = \mu(a) + b\varDelta$. and the equation could be changed into as followed:

$$\frac{\mu(a + \varDelta) - \mu(a)}{\varDelta} = \frac{1}{a^2}\big(2a\mu_{AB} + \varDelta\mu_{BB} - (2a + \varDelta)\mu(a + \varDelta)\big) + b$$

Let μ_1 be the average distance from the increased junction x to the end of this original part, and the approximately distance between part.A and part.B will be written as $\mu_{AB} = \mu_1 + O(\Delta)$, and the following formula could be derived:

$$\frac{d\mu(a)}{da} = \frac{2}{a}\left(\mu_1 - \mu(a)\right) + b \tag{2}$$

For the coefficient b, if the increments are always be inserted in the end of the segment, the cross-section traffic volume from part.A to part.B will be 0. thus the equation could change into as followed:

$$\mu(a) = \frac{2}{3}\left(\mu_1 + \frac{b}{2}a\right) \tag{3}$$

In segment which has length of a referring to fig.24, there is increment in the junction x, thus $\mu_1 = \frac{1}{a}\left(x^2 - ax + \frac{1}{2}a^2\right)$, $b = \frac{2}{a^2}(ax - x^2)$, based on equation.(3), $\mu(a) = \frac{1}{3}a$ could be derived.
Furthermore, in the circle likes fig.28, $\mu_1 = \frac{a}{4}$, $b = \frac{1}{4}$, thus $\mu(a) = \frac{1}{4}a$ based on equation.(3).

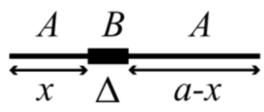


Fig.27 increment in line

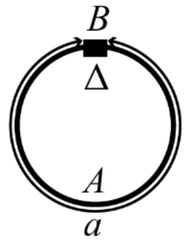


Fig.28 increment in circle

5-1 tree

In tree-network which has k segments, the average distance from the midpoint to the rest parts comes to be $\frac{a}{4}$, and in the whole network, all of the average distance from the midpoints to the rest part of the tree-network go to be as followed:

$$\mu_1 = \bar{d} + \frac{a}{4k} \tag{4}$$

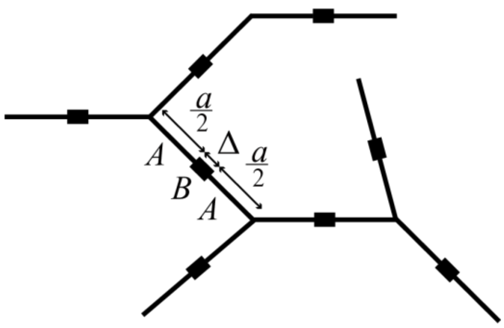


Fig.29 increment in tree-network

Referring to fig.29, there are increments in the middle of every segment. By the same token, let the original part be part.A and the increased part as part.B, the cross-section traffic volume from part.A to part.B(represented as b) will be $\frac{1}{2k}$ when the movement is in the same segments, and will be $\frac{d^2}{a}$ when the movement taking place in different segments. Based on the average distance between the midpoints $\bar{d}$, the cross-section traffic volume from part.A to part.B:

$$b = \frac{1}{a}\left(\bar{d} + \frac{a}{2k}\right) \quad (5)$$

And based on equation.(3), (4), (5), $\mu(a)$could be derived as followed:

$$\mu(a) = \bar{d} + \frac{a}{3k}$$

Which is the same with the result of the general deviation formula.

5-2 lattice networks

By the same token as fig.29, increments are inserted in the middle of every segment, and the original part as part.A while the increased part as part.B(fig.30).

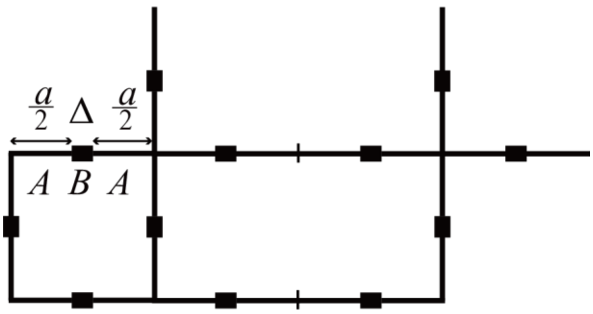


Fig.30 increment in general network

When consider the average distance from every midpoint to the rest of the network, since the exist of antipodal links pair(fig.31), the average distance will decrease by $\frac{a}{4}$.

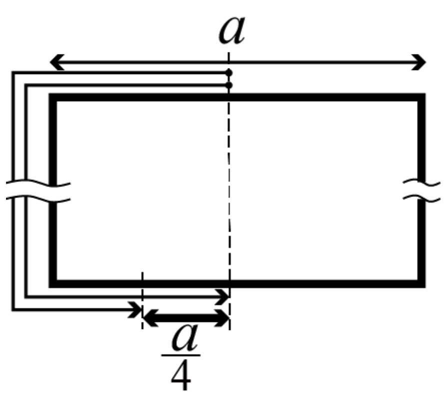


Fig.31　reduced part in average distance

And based on equation.(4) and the average distance between midpoints, the average distance from every midpoint to the rest of the network μ_1 could be derived as followed:

$$\mu_1 = \bar{d} + \frac{(k-\lambda)a}{4k^2} \quad (6)$$

And the the cross-section traffic volume b in the part.A will decrease by $\frac{1}{2}$ (fig.32), thus b will change into as followed:

$$b = \frac{1}{a}\left(\bar{d} + \frac{(k-\lambda)a}{2k^2}\right) \quad (7)$$

Based on equation.(5).

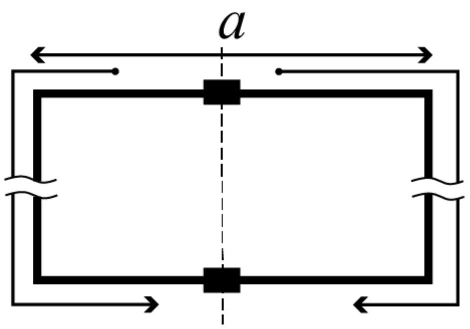


Fig.32 reduced part in cross-section traffic volume

And finally cased on equation.(3), (6), (7),

$$\mu(a) = \bar{d} + \frac{k-\lambda}{3k^2}a$$

could be derived and the result is coincident with the equation.(1)

Conclusion

In this work, the general deviation formula has been derived and it proved the relationship between the discrete average distance and continuous average distance in networks. By analysis, it can be concluded that by calculating the discrete average distances in the network, the continuous average distance can also be derived. And in chapter.5 the general deviation formula is also derived by the using of Crofton's formula. thus the Crofton's formula is applicable in situations in continuous spaces such as line segments, circles, circular rings, ellipses, etc. [8] [9] [10]. And further discussion such as derivation of other moments about distance and probability density function will be conducted in next stage.

References

[1] Y.Ohsawa, T.Koshizuka and O.Kurita(1991): Errors caused by rounded data in two simple facility location problems. *Geographical Analysis*, 23(1), pp.56-73.
[2] T.Koshizuka (1999)：omparison of Road Network Patterns with Respect to Travel Distance and Passing Volume. *Journal of the City Planning Institute of Japan*, 34, pp.763–768．
[3] K.Tamura，T.Koshizuka (2000)：An analytical method to derive the distributions of distances and flow volumes on networks, *Journal of the City Planning Institute of Japan*, 35, pp.1021–1026．
[4] Kendall, M.G. and P.A.P. Moran(1963):*Geometrical Probability*. Charles Griffin and Company Limited.
[5] T.Koshizuka (1986): Distance distribution in an arbitrary region, *Urban Planning Mathematical*. Asakura Publishing Co.
[6] O.Kurita (2004): *Urban-Model Reader*. Kyoritsu Shuppan Co.
[7] Vaughan, R.(1987):*Urban Spatial Traffic Patterns*.Pion.
[8] R.Hori, T.Kobayashii, I.Takahara, Y.Ohsawa（2017）：Average Euclidean distance between two random points in sectors and its applications, *Journal of the City Planning Institute of Japan*, 52,pp.1335-1340.
[9] A.Cho, N.Ishii, I.takahara, Y.Ohsawa（2017）：Average Euclidean distance between two random points in rings, *Journal of the City Planning Institute of Japan*, 52,pp.676-681.
[10] J.Yotsutsuji, A.Kawaguchi, N.Ishii, Y.Ohsawa (2018)：Fan-shaped Sylvester 4 point problem and its application to facility arrangement, *Journal of the City Planning Institute of Japan*,
53, pp.646-652.