Societas Scientiarum Fennica
Commentationes Physico-Mathematicae XXIX 8

ON DISTANCE DISTRIBUTIONS IN NETWORKS

BY

SEPPO MUSTONEN

HELSINKI – HELSINGFORS 1964

Communicated February 17, 1964 by G. ELFVING

CENTRALTRYCKERIET
HELSINGFORS 1964

PREFACE

This subject was suggested to me by Professor LEO TÖRNQVIST, and I wish to express my deep gratitude to him for his valuable advice and inspiring interest in the course of the study.

My gratitude is also extended to Professor GUSTAV ELFVING, who reviewed the original manuscript and who offered many helpful suggestions.

Finally I am indebted to SUOMEN KULTTUURIRAHASTO (The Finnish Cultural Foundation) for its financial support.

Helsinki, April 1964

SEPPO MUSTONEN

CONTENTS

1. INTRODUCTION

Our objective is to study the properties of traffic networks with regard to distance distributions. More precisely, we wish to determine the probability distribution of the distance between two random points in a network.

The basis of our investigations is a certain probability model which has been used previously by Törnqvist in his paper »On Distribution Functions for Quantities Related to Networks» [4]. In the present study we shall more thoroughly describe and justify certain ideas presented by him. We shall also try to develop these ideas both in theory and in practice.

Our method of study differs to some extent from the usual presentation in the theory of graphs, since we shall consider only metric networks. Hence we shall be investigating the network as a connected, non-denumerable point set and not only as a system of distinct and interconnected points. For this reason our terminology differs slightly from that used in the theory of graphs.

The study is divided into five chapters. In the introduction we define the network and some fundamental concepts related to it. Some results from the elementary theory of graphs are also mentioned.

Chapters 2 and 3 are devoted to the metric properties of networks. In Chapter 2 we consider especially questions related to reflection points and reflection sets. Chapter 3 for the most part deals with an important class of networks which we call Törnqvist networks.

In Chapter 4 we concentrate on our main problem, the derivation of the distance distribution under certain conditions. We arrive at an explicit solution for the Törnqvist networks.

In Chapter 5 the problem is extended in a more realistic direction and a solution for general networks is given. At the end of Chapter 5 some potential applications are discussed.

1.1. Definition of the network

Let us take an idealized picture, e.g., of a road network, a system of traffic connections, and call it shortly a *network*.

Definition. A network is a connected metric space consisting of a finite number of simple arcs of finite length, having only end points in common.

In terms of the theory of graphs this is to say that the network is an unoriented, connected graph, to whose edges positive numbers (lengths) are attributed. Such a graph can always be imbedded in a three-dimensional space, and often in a plane.

According to the definition, the network A and each subset $B \subset A$ consisting of arcs has a finite length, which means the sum of the lengths of the arcs concerned. We denote the length of B by $L(B)$ and let especially the length of A be $L(A) = a$.

An immediate consequence of the definition is that two points, α and β, which belong to A can be connected with an arc which belongs to A, has a finite length, and does not intersect itself. We call such an arc a *path* between the points α, β. The points α, β are called the *end points* of the path and they are considered as belonging to the path. Thus a path is a closed set. We will sometimes also say that the path starts from α and ends in β or vice versa.

1.2. Basic concepts

Let α be a point in A. The number of distinct paths starting from α is called the *degree* of α and denoted by n_α.

In the planar network in Fig. 1.1 the degree $n_{\alpha'}$ of α' is equal to 6 and the degree $n_{\alpha''}$ of α'' is equal to 2.

According to the definition of a network, n_α is always finite and ≥ 1. We also notice that the network A possesses only a finite number of points α having $n_\alpha \neq 2$. We call these points *special points.* If $n_\alpha = 1$, α is an *end point* of the network; if $n_\alpha \geq 3$, α is a *junction point.* On the contrary, the points of degree 2 constitute a continuum. These points are called *line points.* In Fig. 1.1, for instance, the points $\varepsilon_1, \varepsilon_2$ and ε_3 are junction points, γ_1 and γ_2 are end points, and α'' is a line point.

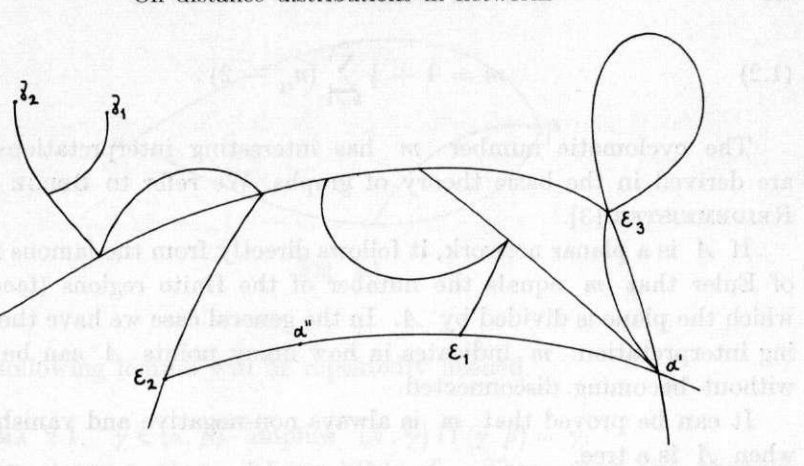


Fig. 1.1

An *edge* is a path between two special points having no special points as inner points. In Fig. 1.1, the path $\varepsilon_1 \alpha'' \varepsilon_2$ is an edge. Between α' and ε_3 there are two edges.

A path whose end points coincide is called a *cycle.* A network containing no cycles is a *tree.* A subset of A to which the definition of the network applies is called a *subnetwork* of A.

1.3. Cyclomatic number

According to the definition of a network, there are only a finite number of special points and edges in a network. Let the numbers of special points and edges be m_0 and m_1, respectively, and let the special points be

$$\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{m_0}$$

The *cyclomatic number* m of a network is defined as

$$m = m_1 - m_0 + 1 \tag{1.1}$$

Since every edge contributes two units to the total degree of the special points, we have

$$m_1 = \tfrac{1}{2} \sum_{k=1}^{m_0} n_{\varepsilon_k}$$

and (1.1) is therefore equivalent to the expression

$$m = 1 + \tfrac{1}{2}\sum_{k=1}^{m_0} (n_{\varepsilon_k} - 2)\,. \tag{1.2}$$

The cyclomatic number m has interesting interpretations which are derived in the basic theory of graphs. We refer to BERGE [1] and REIDEMEISTER [3].

If A is a planar network, it follows directly from the famous formula of Euler that m equals the number of the finite regions (faces) into which the plane is divided by A. In the general case we have the following interpretation: m indicates in how many points A can be cut off without becoming disconnected.

It can be proved that m is always non-negative and vanishes only when A is a tree.

2. METRIC PROPERTIES OF NETWORKS

2.1. Distance in a network

According to the definition of a network, there are only a finite number of paths connecting two points of a network A. Hence there exists a shortest path between any two points, and there are a finite number of such shortest paths.

The shortest paths between two points α, β form a certain subnetwork of A which is denoted by (α, β). The *distance* between the points α, β is defined as the common length $s_{\alpha\beta}$ of these shortest paths. If v is one of the shortest paths between α and β, we usually say briefly that v is an (α, β)-path.

The distance $s_{\alpha\beta}$ is non-negative; it is 0 if and only if α and β coincide. Obviously $s_{\alpha\beta} = s_{\beta\alpha}$. For three arbitrary points α, β, γ, the triangle inequality $s_{\alpha\beta} \leq s_{\alpha\gamma} + s_{\gamma\beta}$ holds with equality if and only if $\gamma \in (\alpha, \beta)$.

The *relative degree* of β with respect to α is the degree of β in the network (α, β); it is denoted by $n_{\beta\alpha}$. In the case $\beta = \alpha$ where (α, β) reduces to the single point β, we define $n_{\beta\beta} = 1$. Usually the shortest path between α and β is unique and the network (α, β) consists of only one (α, β)-path. Then $n_{\alpha\beta} = n_{\beta\alpha} = 1$. There are, however, cases where (α, β) is more complex. In Fig. 2.1, for instance, $n_{\alpha\beta} = 2$ and $n_{\beta\alpha} = 3$.

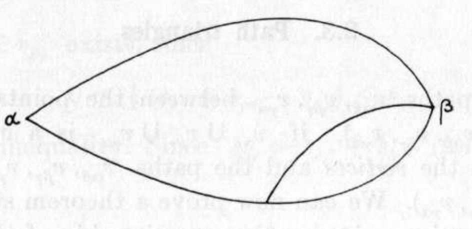


Fig. 2.1

The following lemma will be repeatedly needed.

LEMMA 2.1. $\gamma \in (\alpha, \beta)$ implies $(\alpha, \gamma) \cap (\gamma, \beta) = \gamma$.

PROOF. Suppose that $\delta \in (\alpha, \gamma) \cap (\gamma, \beta)$. Then using the triangle inequality we have

$$\begin{aligned} s_{\alpha\beta} &= s_{\alpha\gamma} + s_{\gamma\beta} = s_{\alpha\delta} + s_{\delta\gamma} + s_{\gamma\delta} + s_{\delta\beta} \\ &= s_{\alpha\delta} + s_{\delta\beta} + 2\, s_{\delta\gamma} \geq s_{\alpha\beta} + 2\, s_{\delta\gamma}\,, \end{aligned}$$

which implies $\delta = \gamma$.

2.2. Reflection points

Let α be an arbitrary point in A. The number of points β, for which $n_{\beta\alpha} > 1$, is finite, since no edge in A can contain more than one line point β satisfying $n_{\beta\alpha} > 1$. As a matter of fact, if an edge would contain two line points β, γ for which $n_{\beta\alpha} > 1$, $n_{\gamma\alpha} > 1$, then we would have at the same time $\beta \in (\alpha, \gamma)$, and hence $s_{\alpha\gamma} = s_{\alpha\beta} + s_{\beta\gamma}$, and $\gamma \in (\alpha, \beta)$, and hence $s_{\alpha\beta} = s_{\alpha\gamma} + s_{\gamma\beta}$. But this would imply $s_{\beta\gamma} = 0$ and thus $\beta = \gamma$.

The points β for which $n_{\beta\alpha} > 1$ are called the *reflection points* of α in A. The reflection points will play a very important part in the following, due to the fact that they are points in which the distance $s_{\alpha\beta}$ attains local maximum values. This fact will later be brought to the fore in Theorem 3.1.

In the following we shall present some basic theorems concerning reflection points.

2.3. Path triangles

The shortest paths $v_{\alpha\beta}, v_{\beta\gamma}, v_{\gamma\alpha}$ between the points α, β, γ form a *path triangle* $\Delta(v_{\alpha\beta}, v_{\beta\gamma}, v_{\gamma\alpha})$, if $v_{\alpha\beta} \cup v_{\beta\gamma} \cup v_{\gamma\alpha}$ is a cycle. The points α, β, γ are called the *vertices* and the paths $v_{\alpha\beta}, v_{\beta\gamma}, v_{\gamma\alpha}$ are called the *sides* of $\Delta(v_{\alpha\beta}, v_{\beta\gamma}, v_{\gamma\alpha})$. We can now prove a theorem stating that every vertex has a reflection point on the opposite side of the path triangle.

THEOREM 2.2. The vertex α of $\Delta(v_{\alpha\beta}, v_{\beta\gamma}, v_{\gamma\alpha})$ has, on the opposite side $v_{\beta\gamma}$, at least one reflection point α' fulfilling $\gamma \in (\alpha, \alpha')$.

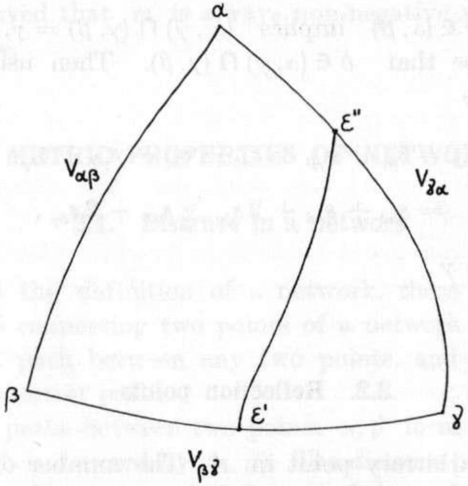


Fig. 2.2

PROOF. We consider the paths between α and an arbitrary point $\delta \in v_{\beta\gamma}$ and denote

$$\bar{s}_{\delta\alpha} = \min(s_{\delta\beta} + s_{\beta\alpha}, s_{\delta\gamma} + s_{\gamma\alpha}).$$

Thus $s_{\delta\alpha} \leq \bar{s}_{\delta\alpha}$. The following proof is divided into two parts depending on whether $s_{\delta\alpha} = \bar{s}_{\delta\alpha}$ for all $\delta \in v_{\beta\gamma}$ or not.

1° Let us first suppose that $s_{\delta\alpha} = \bar{s}_{\delta\alpha}$ for all $\delta \in v_{\beta\gamma}$: in other words, at least one of the (δ, α)-paths goes either through β or γ. Consider the point $\delta \in v_{\beta\gamma}$ for which

$$s_{\delta\beta} = \tfrac{1}{2}(s_{\beta\gamma} + s_{\gamma\alpha} - s_{\alpha\beta}). \tag{2.1}$$

Such a point $\delta \in v_{\beta\gamma}$ exists, since

$$0 \leq \tfrac{1}{2}(s_{\beta\gamma} + s_{\gamma\alpha} - s_{\alpha\beta}) \leq s_{\beta\gamma}$$

by the triangle inequality. Since $s_{\beta\gamma} = s_{\delta\beta} + s_{\delta\gamma}$, (2.1) can be written in the form

$$s_{\delta\beta} + s_{\beta\alpha} = s_{\delta\gamma} + s_{\gamma\alpha}$$

from which it follows that

$$s_{\delta\alpha} = \bar{s}_{\delta\alpha} = \min(s_{\delta\beta} + s_{\beta\alpha}, s_{\delta\gamma} + s_{\gamma\alpha})$$
$$= s_{\delta\beta} + s_{\beta\alpha} = s_{\delta\gamma} + s_{\gamma\alpha}.$$

Hence δ is a reflection point of α and $\gamma \in (\alpha, \delta)$ so that δ can be chosen for the point α'.

2° We now suppose that there exists a point $\delta \in v_{\beta\gamma}$ for which $s_{\delta\alpha} < \bar{s}_{\delta\alpha}$. Thus the (δ, α)-paths do not pass through β nor through γ. Since $v_{\alpha\beta} \cup v_{\beta\gamma} \cup v_{\gamma\alpha}$ is a cycle and therefore $\alpha \notin v_{\beta\gamma}$, we conclude that there exists a junction point $\varepsilon' \in v_{\beta\gamma}$ belonging to (δ, α) and such that the (ε', α)-paths have no common points with $v_{\beta\gamma}$ except ε'. If $v_{\beta\gamma}$ has several junction points fulfilling this condition, we take for ε' the point which is nearest to γ.

Let v be one of the (ε', α)-paths and let ε'' be the nearest to ε' junction point in which v meets $v_{\gamma\alpha}$. If such a point ε'' does not exist (i.e., v meets $v_{\alpha\beta}$), we take $\varepsilon'' = \alpha$. Then we have a new path triangle with vertices $\varepsilon'', \varepsilon', \gamma$ and it can be shown, as in the first part of the proof, that ε'' has a reflection point $\alpha' \in (\varepsilon', \gamma) \cap v_{\beta\gamma}$ fulfilling $\gamma \in (\varepsilon'', \alpha')$. But since $\varepsilon'' \in (\alpha, \varepsilon')$, $\varepsilon'' \in (\alpha, \gamma)$, and, according to the definition of ε', either $\varepsilon' \in (\alpha, \alpha')$ or $\gamma \in (\alpha, \alpha')$, we must have $\varepsilon'' \in (\alpha, \alpha')$. Since α' is a reflection point of ε'' fulfilling $\gamma \in (\varepsilon'', \alpha')$ and $\varepsilon'' \in (\alpha, \alpha')$, it follows that α' is also a reflection point of α and $\gamma \in (\alpha, \alpha')$. Hence α' satisfies the requirements of the theorem.

2.4. Classification of reflection points

We denote by $T(\alpha)$ the finite set of all reflection points of the point α. This set will be divided into two subsets. We say that β is an *essential* reflection point of α, if $\beta \in T(\alpha)$ and if the network (α, β) contains a cycle to which both α and β belong. Let $T_1(\alpha)$ be the set of the essential reflection points of α. It is immediately seen that $\beta \in T_1(\alpha)$

implies $\alpha \in T_1(\beta)$. The reflection points of α which are not essential reflection points are called *unessential*. They constitute a set $T_2(\alpha)$. Thus $T(\alpha)$ is the union of the disjoint sets $T_1(\alpha)$ and $T_2(\alpha)$.

Intuitively, if $\beta \in T_1(\alpha)$, a short displacement of α along a certain path will always cause a similar displacement of β; see Fig. 2.3 a. On the other hand, if $\beta \in T_2(\alpha)$, the point β remains fixed although α is displaced; see Fig. 2.3 b. The objective of the two following theorems is to express these properties of reflection points in a more accurate form. Let us first consider the unessential reflection points.

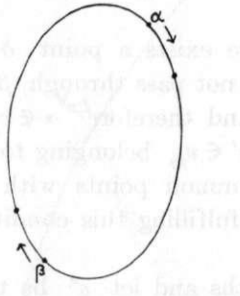


Fig. 2.3 a

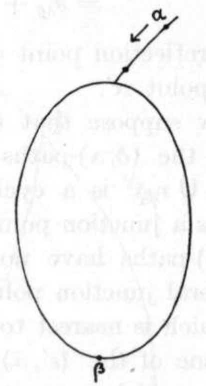


Fig. 2.3 b

THEOREM 2.3. If β is an unessential reflection point of α, then there exists at least one junction point in (α, β) for which β is an essential reflection point.

PROOF. Since $\beta \in T(\alpha)$ and therefore $n_{\beta\alpha} > 1$, there exist at least two (α, β)-paths, starting from β in different directions. Let v_1 and v_2 be such paths and let γ be their common point nearest to β. Such a point γ exists, since v_1 and v_2 meet at least in α. The point γ cannot, however, coincide with α, for in this case $v_1 \cup v_2$ would be a cycle and $\beta \in T_1(\alpha)$, contrary to our assumption.

Since $n_{\gamma\beta} \geq 2$, and since γ is, by Lemma 2.1, the only common point of (α, γ) and (γ, β), we have $n_\gamma \geq 3$. Hence γ is a junction point and belongs to (α, β). Furthermore, $\beta \in T_1(\gamma)$, since the parts of v_1 and v_2 bounded by β and γ form a cycle.

In order to establish a corresponding theorem for essential reflection points, we must first define the concept of a *basic cycle*.

DEFINITION. *A cycle D is a basic cycle, if between any two points α, β of D there is a shortest path which belongs to D.*

REMARK. It would be possible to restrict the condition of the definition to the junction points of the cycle. Actually, if a cycle C contains two points α, β such that no (α, β)-path belongs entirely to C, then C has also two junction points the shortest paths between which have no points in common with C except their end points. This fact can be established by inspecting the end points of the common parts of some (α, β)-path and C.

An immediate consequence of the definition of a basic cycle is that every point on a basic cycle D has exactly one essential reflection point on the same basic cycle D at a distance of $\frac{1}{2} L(D)$. In addition, if α and β are two arbitrary points of D, the distance between their reflection points on D is the same as the distance $s_{\alpha\beta}$ between α and β.

The following theorem is a generalization of these statements.

THEOREM 2.4. Let β be an essential reflection point of α and let C be a cycle belonging to the network (α, β) and containing α and β. (The existence of such a cycle follows from the definition of an essential reflection point.)

If C is not a basic cycle, then it contains two paths v_α (with end points α', α'' and containing α), v_β (with end points β', β'' and containing β) possessing the following properties:

1° $L(v_\alpha) = s_{\alpha'\alpha''}$, $L(v_\beta) = s_{\beta'\beta''}$.
2° $L(v_\alpha) = L(v_\beta)$.
3° If $\bar{\alpha} \in v_\alpha$, there exists a $\bar{\beta} \in v_\beta$ such that $\bar{\beta} \in T_1(\bar{\alpha})$ and $s_{\bar{\alpha}\bar{\beta}} = s_{\alpha\beta} = \frac{1}{2} L(C)$.
4° The end points of v_α and v_β are essential reflection points of some junction points on C.

PROOF. The points α and β divide C into two paths v' and v'' both of which are (α, β)-paths. Since, by assumption, C is not a basic cycle, it contains at least one pair of junction points $\varepsilon', \varepsilon''$ the shortest paths between which have no points in common with C except their end points. This is clear from the remark made in connection with the definition of a basic cycle. Let u be one of the $(\varepsilon', \varepsilon'')$-paths.

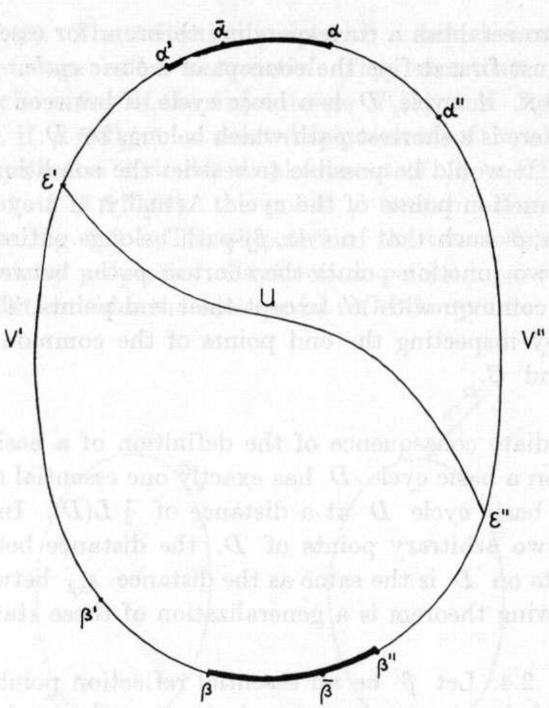


Fig. 2.4

Both of the points $\varepsilon', \varepsilon''$ cannot be on v', since v' is an (α, β)-path. For the same reason both of them cannot be on v''. Then let us suppose that $\varepsilon' \in v'$ and $\varepsilon'' \in v''$. Let u' be the part of v' bounded by α and ε' and let u'' be the part of v'' bounded by α and ε''. Since $u \cup u' \cup u''$ is a cycle formed by the shortest paths between the points $\alpha, \varepsilon', \varepsilon''$, it can be seen that $\Delta(u, u', u'')$ is a path triangle. By applying Theorem 2.2, we find that ε'' has on u' a reflection point α' fulfilling $\varepsilon' \in (\varepsilon'', \alpha')$. If there are several such reflection points, we choose the one nearest to α for α'.

We now consider all pairs of points $\varepsilon', \varepsilon''$ for which the previous considerations are valid and pick the pair of points $\varepsilon', \varepsilon''$ that gives the α' nearest to α. We are going to show that α' is then an essential reflection point of ε''. Denote by u'_α the part of v' bounded by α and α'. Let $\bar\alpha$ be an inner point of u'_α (i.e. $\bar\alpha \neq \alpha, \alpha'$). We will now prove that $\alpha' \notin (\varepsilon'', \bar\alpha)$.

Indeed, suppose that $\alpha' \in (\varepsilon'', \bar\alpha)$ and hence also $\varepsilon' \in (\varepsilon'', \bar\alpha)$, as a consequence of $\varepsilon' \in (\varepsilon'', \alpha')$. The paths $u'', v' \cap (\alpha, \bar\alpha)$, and $u \cup (v' \cap (\varepsilon', \bar\alpha))$ would then form a path triangle with vertices $\varepsilon'', \bar\alpha, \alpha$. Hence, by Theorem 2.2, ε'' would have on v', between the points $\bar\alpha, \alpha$, a reflection point γ for which $\bar\alpha \in (\varepsilon'', \gamma)$ and hence also $\varepsilon' \in (\varepsilon'', \gamma)$, since $\varepsilon' \in (\varepsilon'', \bar\alpha)$. This, however, contradicts our assumption that $v' \cap (\alpha, \bar\alpha)$ does not contain any reflection point γ of ε'' satisfying $\varepsilon' \in (\varepsilon'', \gamma)$. Thus it has been proved that $\alpha' \notin (\varepsilon'', \bar\alpha)$.

We shall now prove that $\alpha \in (\varepsilon'', \bar\alpha)$. The proof is indirect. Suppose that $\alpha \notin (\varepsilon'', \bar\alpha)$. Let v be an $(\varepsilon'', \bar\alpha)$-path. Since, by assumption, $\alpha \notin (\varepsilon'', \bar\alpha)$ and, by the previous argument, $\alpha' \notin (\varepsilon'', \bar\alpha)$, there would have to exist on v', between α' and α, a junction point $\bar\varepsilon'$ in which v takes off from v'. Since $\varepsilon'' \in v''$, there must be another junction point $\bar\varepsilon'' \in v''$ in which v joins v''.

Let $\bar u$ be the part of v bounded by the points $\bar\varepsilon'$ and $\bar\varepsilon''$. We may assume that $\bar u$ has no points in common with C except $\bar\varepsilon'$ and $\bar\varepsilon''$. Since now $\Delta(\bar u, v' \cap (\bar\varepsilon', \alpha), v'' \cap (\bar\varepsilon'', \alpha))$ is a path triangle with vertices $\alpha, \bar\varepsilon', \bar\varepsilon''$, the vertex $\bar\varepsilon''$ would then have, by Theorem 2.2, on v' between $\bar\varepsilon'$ and α a reflection point γ for which $\bar\varepsilon' \in (\bar\varepsilon'', \gamma)$. But this contradicts our assumption that α' is, of all points of this type, the one nearest to α.

Hence it has been proved that $\alpha \in (\varepsilon'', \bar\alpha)$ for every inner point $\bar\alpha$ of u'_α; but then also $\alpha \in (\varepsilon'', \alpha')$, because α' is an end point of u'_α.

Since $\varepsilon' \in (\varepsilon'', \alpha')$ and $\alpha \in (\varepsilon'', \alpha')$, the cycle $u \cup u' \cup u''$ belongs to the network (ε'', α') and it has been shown that $\alpha' \in T_1(\varepsilon'')$.

Next we consider an arbitrary inner point $\bar\alpha$ of u'_α and denote by $\bar\beta$ that point on v'' for which $s_{\beta\bar\beta} = s_{\alpha\bar\alpha}$. We wish to prove that $C \subset (\bar\alpha, \bar\beta)$.

Again, the proof is indirect. Assume that $C \not\subset (\bar\alpha, \bar\beta)$. This is to say that $\alpha \notin (\bar\alpha, \bar\beta)$. Let v be one of the $(\bar\alpha, \bar\beta)$-paths. Since $\alpha \notin (\bar\alpha, \bar\beta)$, there must be junction points $\bar\varepsilon' \in v'$ and $\bar\varepsilon'' \in v''$ in which v takes off from the cycle C and such that v has no points in common with C between $\bar\varepsilon'$ and $\bar\varepsilon''$.

If $\bar\alpha \in (\bar\varepsilon', \alpha)$, we have a path triangle with vertices $\bar\alpha, \alpha, \bar\varepsilon''$ and if $\bar\varepsilon' \in (\bar\alpha, \alpha)$, we have a path triangle with vertices $\bar\varepsilon', \alpha, \bar\varepsilon''$. In both cases it is seen immediately, by Theorem 2.2, that $\bar\varepsilon''$ has on v' between $\bar\alpha$ and α a reflection point γ for which $\bar\varepsilon' \in (\bar\varepsilon'', \gamma)$. This contradicts the assumptions about ε' and ε''. Hence $C \subset (\bar\alpha, \bar\beta)$. This is also to say that $s_{\bar\alpha\bar\beta} = \frac{1}{2} L(C)$ and $\bar\beta \in T_1(\bar\alpha)$. If we let $\bar\alpha$ approach α', then $\bar\beta$ will approach a point β'' for which, by continuity, the relations $\beta'' \in T_1(\alpha')$, $s_{\beta\beta''} = s_{\alpha\alpha'}$, $C \subset (\alpha', \beta'')$ will still hold. The point β'' is an essential reflection point of ε'. In fact, since $\alpha' \in T_1(\varepsilon'')$ and $C \subset (\alpha', \beta'')$, we have

$$\tfrac{1}{2} L(C) = s_{\alpha'\varepsilon'} + s_{\varepsilon'\varepsilon''} + s_{\varepsilon''\beta''}$$

$$= s_{\alpha'\varepsilon'} + s_{\varepsilon'\beta} + s_{\beta\beta''}$$

or

$$s_{\varepsilon'\varepsilon''} + s_{\varepsilon''\beta''} = s_{\varepsilon'\beta} + s_{\beta\beta''}$$

which implies that there exists a cycle belonging to (ε', β'') and containing ε' and β''. Thus $\beta'' \in T_1(\varepsilon')$.

A similar argument can be carried out for v'' as has now been performed for v'. As a result we find points $\alpha'' \in v''$, $\beta' \in v'$, and a path $u''_\alpha \subset v''$, with properties similar to those of α', β'', and u'_α, respectively. If we denote $v_\alpha = u'_\alpha \cup u''_\alpha$ and define v_β in a similar manner, the assertions 1°—4° of Theorem 2.4 can be easily verified:

1° follows for v_α from $C \subset (\alpha', \beta'')$.

2° is true, since it has been shown that $s_{\alpha\alpha'} = s_{\beta\beta''}$, and similarly, $s_{\alpha\alpha''} = s_{\beta\beta'}$.

3° has been proved above for $\bar{\alpha} \in u'_\alpha$.

4° has been proved above for α', β''.

Remark 1. The paths v_α, v_β may occasionally reduce to the points α, β.

Remark 2. If C is a basic cycle, we may take $v_\alpha = v_\beta = C$; assertions 2° and 3° are then valid also in this case.

2.5. Reflection sets

We shall generalize the reflection point concept extending it so as to apply to the subnetworks of A. Let B be an arbitrary subnetwork. The *reflection set* $T(B)$ of B is defined as

$$T(B) = \bigcup_{\alpha \in B} T(\alpha) .$$

The essential and unessential reflection sets are defined in a similar manner by

$$T_1(B) = \bigcup_{\alpha \in B} T_1(\alpha) ,$$

$$T_2(B) = \bigcup_{\alpha \in B} T_2(\alpha)$$

and hence

$$T(B) = T_1(B) \cup T_2(B) .$$

If B consists of a single point, the sets $T_1(B)$ and $T_2(B)$ have no points in common. If B is a non-degenerate subnetwork, this need not be so. Those cases in which

$$T_1(B) \cap T_2(B) = \emptyset , \tag{2.2}$$

offer a particular interest. We say that B is *completely reflective*, if it satisfies (2.2).

The nature of a completely reflective subnetwork is illustrated in Fig. 2.5; the subnetwork (path) v is completely reflective whereas v' is not. The fact that v' is not completely reflective is due to the degenerating of one of the components of $T_1(v')$. Certain reflection points of the points between α_1 and α_2 coincide in the single point γ which is an unessential reflection point of all of them. Since, however, $\gamma \in T_1(\alpha_1)$, we have $T_1(v') \cap T_2(v') = \gamma \neq \emptyset$.

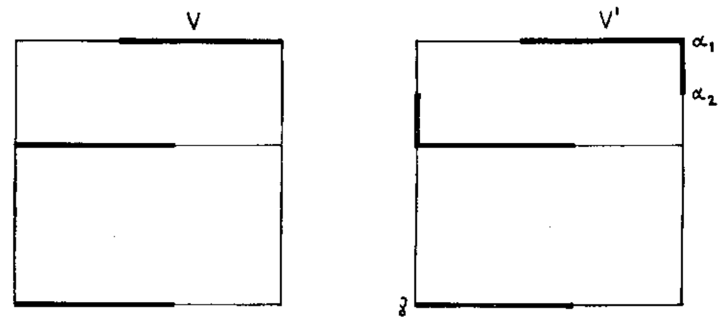


Fig. 2.5

In this example the unessential reflection set was empty or consisted of a single point. As a matter of fact, $T_2(B)$ is always a finite set since, according to Theorem 2.3, $T_2(B)$ consists of reflection points of junction points.

We shall now study under what conditions a path in A is completely reflective.

THEOREM 2.5. A path $v \in A$ that does not (even as end points) contain junction points, or reflection points of junction points, is completely reflective. $T_1(v)$ then consists of a finite number of paths all having the same length $L(v)$ but no common points.

PROOF. By assumption, v is part of an edge. To prove the first part of the theorem, let us first suppose that no point of v has any essential reflection points. Since $T_1(v) = \emptyset$ and therefore also $T_1(v) \cap T_2(v) = \emptyset$, it is seen that v is completely reflective. We then suppose that a point $\alpha \in v$ has an essential reflection point β. Using Theorem 2.4 it can be verified that β cannot be an unessential reflection point of any point of v. This proves that v is completely reflective. In fact, we observe that since v does not contain junction points and reflection points of junction points, v is, by Theorem 2.4, 4°, a part of the path v_α defined in that theorem.

Since $\beta \in T_1(\alpha)$, β cannot be a reflection point of a point $\bar{\alpha} \in v$ different from α. This follows from Theorem 2.4, 3° according to which $\bar{\alpha}$ has an essential reflection point $\bar{\beta}$ such that $\beta \in (\bar{\alpha}, \bar{\beta})$. In order for β to be a reflection point of $\bar{\alpha}$, β would have to be a junction point, since it is, by Lemma 2.1, the only common point of $(\bar{\alpha}, \beta)$ and $(\beta, \bar{\beta})$. This is impossible, since a junction point β would have a reflection point $\alpha \in v$. Thus the first part of the theorem is proved.

The second part follows directly from the previous considerations by which it can be seen that $T_1(v)$ consists of a finite number of paths, whose end points are the essential reflection points of the end points of v and each one of which has the length $L(v)$. These paths, the components of $T_1(v)$, have no common points, since if two of them had a common point, it would mean that such a common point would have at least two different reflection points on v, which is impossible (cf. 2.2).

A direct consequence of Theorem 2.5 is

COROLLARY 2.5.1. For a path $v \in A$ which does not contain junction points, or reflection points of junction points as *inner* points, $T(v)$ consists of a finite number of distinct points and paths all having the same length $L(v)$ but no common inner points.

The general solution of our main problem, viz., the determination of the distance distribution, is based on this corollary. A path v to which Corollary 2.5.1 applies may be called completely reflective in a wide sense.

2.6. Nonsingular networks

A network A is *nonsingular*, if the relative degree $n_{\beta\alpha}$ (cf. 2.1) is at most 2 for all points α, β in A. Networks appearing in practice are usually nonsingular, since the condition $n_{\beta\alpha} \geq 3$ requires that there be at least three paths of equal length between α and β.

Probably the most important class of nonsingular networks consists of those in which the shortest path between any two junction points is unique.

2.7. The function $n_\alpha(s)$

Let α be an arbitrary fixed point in A. We introduce the non-negative, integer-valued function defined by

$$n_\alpha(s) = \text{number of points } \beta \in A \text{ for which } s_{\alpha\beta} = s.$$

This function is always bounded, since no edge can contain more than two points at a given distance s from α. We now try to find an explicit expression for $n_\alpha(s)$.

The jump (if any) of $n_\alpha(s)$ at a certain distance s is built up of the »component jumps» contributed by all points β for which $s_{\alpha\beta} = s$. The jump contributed by such a point is

$$(n_\beta - n_{\beta\alpha}) - n_{\beta\alpha} = n_\beta - 2\,n_{\beta\alpha}\,.$$

$$n_\beta - 2\,n_{\beta\alpha} = 5 - 4 = 1$$

Fig. 2.6

This expression can be different from zero only if β belongs either to $T(\alpha)$, the set of the reflection points of α, or to $E = \{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{m_0}\}$, the set of the special points of A. Thus $n_\alpha(s)$ has only a finite number of discontinuities and these can occur only at the points

$$(2.3) \qquad \begin{array}{l} s = 0 \ , \\ s = s_{\alpha\beta} \, , \quad \beta \in T(\alpha) \cup E \, . \end{array}$$

Between two points of discontinuity $n_\alpha(s)$ is constant. Thus $n_\alpha(s)$ is a non-negative, integer-valued step function.

The definition of $n_\alpha(s)$ is ambiguous for the values (2.3). We amend the definition by taking the limit on the right as the value of $n_\alpha(s)$ at the points (2.3).

In order to establish a formula for $n_\alpha(s)$, we introduce the auxliary function

$$\varphi(x) = \begin{cases} 1 \, , & \text{if } x \geq 0 \\ 0 \, , & \text{if } x < 0 \, . \end{cases}$$

Observing that $n_\alpha(0) = n_\alpha$ and $n_{\alpha\alpha} = 1$, by the definition in 2.1, we then obtain

$$(2.4) \qquad n_\alpha(s) = 2\,\varphi(s) + \sum_\beta (n_\beta - 2\,n_{\beta\alpha})\,\varphi(s - s_{\alpha\beta}) \, ,$$

where the sum is taken over all $\beta \in T(\alpha) \cup E$. In fact, in all other points the terms of the sum vanish.

We may modify (2.4) a little using the identity

$$n_\beta - 2\,n_{\beta\alpha} = (n_\beta - 2) - 2(n_{\beta\alpha} - 1)$$

and splitting up (2.4) correspondingly into

$$n_\alpha(s) = 2\,\varphi(s) + \sum_\beta (n_\beta - 2)\,\varphi(s - s_{\alpha\beta}) - 2 \sum_\beta (n_{\beta\alpha} - 1)\,\varphi(s - s_{\alpha\beta}) \, ,$$

where $\beta \in T(\alpha) \cup E$ as before. However, since $n_\beta - 2 = 0$, if $\beta \notin E$, and $n_{\beta\alpha} - 1 = 0$, if $\beta \notin T(\alpha)$, we have

$$(2.5) \quad n_\alpha(s) = 2\,\varphi(s) + \sum_{k=1}^{m_0} (n_{\varepsilon_k} - 2)\,\varphi(s - s_{\alpha\varepsilon_k}) - 2 \sum_{\beta \in T(\alpha)} (n_{\beta\alpha} - 1)\,\varphi(s - s_{\alpha\beta}) \, .$$

2.8. The number of reflection points

We apply (2.5) to evaluate the number of reflection points of α. According to the definitions of a network and of $n_\alpha(s)$, the function $n_\alpha(s)$ vanishes, when s becomes large enough. More precisely, we know that $n_\alpha(s) = 0$, if $s \geq \hat{s}_\alpha = \max\limits_\beta s_{\alpha\beta}$, where $\beta \in T(\alpha) \cup E$. If $s \geq \hat{s}_\alpha$, we have $\varphi(s - s_{\alpha\beta}) = 1$ for all β, and thus

$$0 = 2 + \sum_{k=1}^{m_0} (n_{\varepsilon_k} - 2) - 2 \sum_{\beta \in T(\alpha)} (n_{\beta\alpha} - 1) \, .$$

Hence, by (1.2),

$$(2.6) \qquad \sum_{\beta \in T(\alpha)} (n_{\beta\alpha} - 1) = 1 + \tfrac{1}{2} \sum_{k=1}^{m_0} (n_{\varepsilon_k} - 2) = m \, .$$

This proves

THEOREM 2.7. In a nonsingular network the number of reflection points of any point equals the cyclomatic number m of the network.

This theorem can be formally extended so as to apply to all networks by interpreting $n_{\beta\alpha} - 1$ as the *order* of the reflection point β; one may imagine that $n_{\beta\alpha} - 1$ distinct, and simple, reflection points have merged into one.

3. EXTREME POINTS

3.1. Definition

Conceptually closely related to the reflection points are the *extreme points* which are defined as follows:

Let B_1 and B_2 be subnetworks of A. We say that a point β_2 belonging to B_2 is an extreme point of B_1 on B_2 if for some point β_1 belonging to B_1

$$s_{\beta_1\beta_2} = \sup_{\substack{\alpha_1 \in B_1 \\ \alpha_2 \in B_2}} s_{\alpha_1\alpha_2} \, .$$

Conversely, β_1 is then also an extreme point of B_2 on B_1.

The set of the extreme points of B_1 on B_2 is denoted by $A(B_1, B_2)$ and the extreme distance $s_{\beta_1\beta_2}$ by $s(B_1, B_2)$. For instance, in the network shown in Fig. 3.1 and consisting of the basic cycles D_1 and D_2, we have $A(D_1, D_2) = (\beta_2, \beta_2')$, $A(D_2, D_1) = (\beta_1, \beta_1')$ and $s(D_1, D_2) = s_{\beta_1'\beta_2'} = s_{\beta_1\beta_2}$.

The following theorem illustrates the nature of the extreme points and shows their relationship to the reflection points.

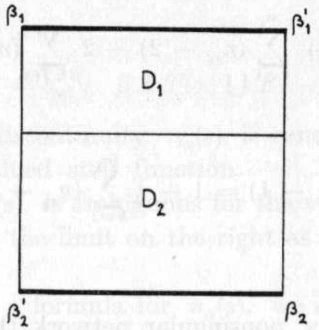


Fig. 3.1

THEOREM 3.1. Let α be an arbitrary point in A and let B be a subnetwork of A. Then $A(\alpha, B)$ consists of distinct points in B each of which is either an end point of B or a reflection point of α.

REMARK. If B has no end points, then $A(\alpha, B) \subset T(\alpha)$.

PROOF. Let β be an extreme point of α on B. We make the assumption that β is neither an end point of B nor a reflection point of α, and show that this leads to a contradiction to the assumption that β is an extreme point.

According to the antithesis, we have $n_\beta \geq 2$ in B and $n_{\beta\alpha} = 1$. This is to say that there is at least one edge or part of an edge of A in B, say v, starting from β but not belonging to any (α, β)-path. We denote by γ the other end point of v. Since $v \not\subset (\alpha, \beta)$, we have

$$s_{\alpha\beta} < s_{\alpha\gamma} + L(v)\,. \tag{3.1}$$

From any inner point δ of v, the shortest path to α must pass through γ, since otherwise $s_{\alpha\delta} > s_{\alpha\beta}$, against the extreme point property of β. Thus $s_{\alpha\beta} > s_{\alpha\gamma} + s_{\gamma\delta}$, which contradicts (3.1) when δ is close enough to β.

3.2. Extreme points on basic cycles

We shall now study the extreme points on basic cycles more closely. In 2.4 we defined a basic cycle as a cycle D which satisfies the following condition: to any two points α, β of D there is an (α, β)-path which belongs to D.

We first consider the case that B_1 is a point α whereas B_2 is a basic cycle D. According to Theorem 3.1, $A(\alpha, D)$ always consists of reflection points of α on D; in fact we shall prove more:

THEOREM 3.2. If β is a reflection point of α, then there exists a basic cycle D containing β, and we have $A(\alpha, D) = \beta$ (i.e. β is the only extreme point of α on D).

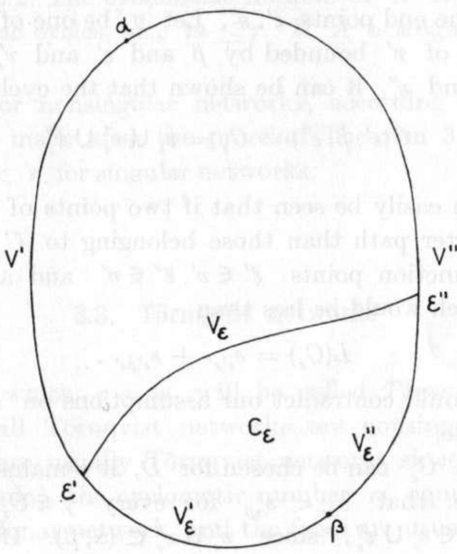


Fig. 3.2

PROOF. We first assume that $\beta \in T_1(\alpha)$. By the definition of an essential reflection point, there exists a cycle C belonging to (α, β) and containing α and β. The points α, β divide C into two (α, β)-paths v', v''. Let $\varepsilon' \in v'$ and $\varepsilon'' \in v''$ be junction points and denote

$$s_{\varepsilon'\beta\varepsilon''} = s_{\varepsilon'\beta} + s_{\beta\varepsilon''}\,,$$

$$s_{\varepsilon'\alpha\varepsilon''} = s_{\varepsilon'\alpha} + s_{\alpha\varepsilon''}\,.$$

We shall distinguish two cases.

1° For all $\varepsilon' \in v'$, $\varepsilon'' \in v''$, the equation

$$s_{\varepsilon'\varepsilon''} = \min\,(s_{\varepsilon'\alpha\varepsilon''}\,, s_{\varepsilon'\beta\varepsilon''})$$

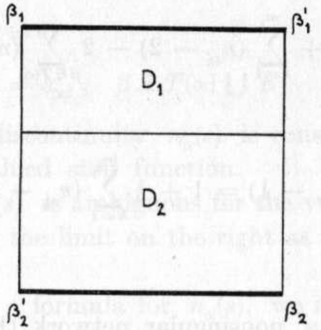


Fig. 3.1

THEOREM 3.1. Let α be an arbitrary point in A and let B be a subnetwork of A. Then $A(\alpha, B)$ consists of distinct points in B each of which is either an end point of B or a reflection point of α.

REMARK. If B has no end points, then $A(\alpha, B) \subset T(\alpha)$.

PROOF. Let β be an extreme point of α on B. We make the assumption that β is neither an end point of B nor a reflection point of α, and show that this leads to a contradiction to the assumption that β is an extreme point.

According to the antithesis, we have $n_\beta \geq 2$ in B and $n_{\beta\alpha} = 1$. This is to say that there is at least one edge or part of an edge of A in B, say v, starting from β but not belonging to any (α, β)-path. We denote by γ the other end point of v. Since $v \not\subset (\alpha, \beta)$, we have

$$s_{\alpha\beta} < s_{\alpha\gamma} + L(v) \,. \tag{3.1}$$

From any inner point δ of v, the shortest path to α must pass through γ, since otherwise $s_{\alpha\delta} > s_{\alpha\beta}$, against the extreme point property of β. Thus $s_{\alpha\beta} > s_{\alpha\gamma} + s_{\gamma\delta}$, which contradicts (3.1) when δ is close enough to β.

3.2. Extreme points on basic cycles

We shall now study the extreme points on basic cycles more closely. In 2.4 we defined a basic cycle as a cycle D which satisfies the following condition: to any two points α, β of D there is an (α, β)-path which belongs to D.

We first consider the case that B_1 is a point α whereas B_2 is a basic cycle D. According to Theorem 3.1, $A(\alpha, D)$ always consists of reflection points of α on D; in fact we shall prove more:

THEOREM 3.2. If β is a reflection point of α, then there exists a basic cycle D containing β, and we have $A(\alpha, D) = \beta$ (i.e. β is the only extreme point of α on D).

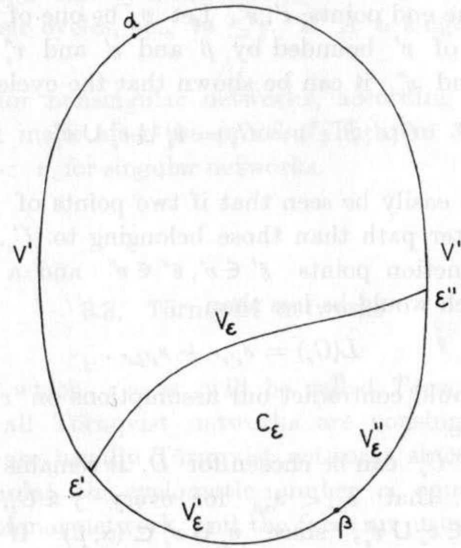


Fig. 3.2

PROOF. We first assume that $\beta \in T_1(\alpha)$. By the definition of an essential reflection point, there exists a cycle C belonging to (α, β) and containing α and β. The points α, β divide C into two (α, β)-paths v', v''. Let $\varepsilon' \in v'$ and $\varepsilon'' \in v''$ be junction points and denote

$$s_{\varepsilon'\beta\varepsilon''} = s_{\varepsilon'\beta} + s_{\beta\varepsilon''} \,,$$

$$s_{\varepsilon'\alpha\varepsilon''} = s_{\varepsilon'\alpha} + s_{\alpha\varepsilon''} \,.$$

We shall distinguish two cases.

1° For all $\varepsilon' \in v'$, $\varepsilon'' \in v''$, the equation

$$s_{\varepsilon'\varepsilon''} = \min\,(s_{\varepsilon'\alpha\varepsilon''}\,, s_{\varepsilon'\beta\varepsilon''})$$

is valid. Then, by the definition of a basic cycle, C is the basic cycle whose existence is asserted in the theorem.

2° There exists a pair of junction points $\varepsilon' \in v'$, $\varepsilon'' \in v''$ for which

$$s_{\varepsilon'\varepsilon''} < \min (s_{\varepsilon'\alpha\varepsilon''}, s_{\varepsilon'\beta\varepsilon''}) .$$

If there are several pairs of points fulfilling this relation we shall examine the pair of points $\varepsilon', \varepsilon''$ for which $s_{\varepsilon'\varepsilon''} + s_{\varepsilon'\beta\varepsilon''}$ is minimum. This pair can be chosen so that none of the $(\varepsilon', \varepsilon'')$-paths has any common points with C except the end points $\varepsilon', \varepsilon''$. Let v_ε be one of the $(\varepsilon', \varepsilon'')$-paths. If v'_ε is the part of v' bounded by β and ε' and v''_ε is the part of v'' bounded by β and ε'', it can be shown that the cycle

$$C(\varepsilon', \beta, \varepsilon'') = C_\varepsilon = v_\varepsilon \cup v'_\varepsilon \cup v''_\varepsilon$$

is a basic cycle.

In fact, it can easily be seen that if two points of C_ε could be connected by a shorter path than those belonging to C_ε, there would be a new pair of junction points $\bar\varepsilon' \in v', \bar\varepsilon'' \in v''$ and a cycle $C(\bar\varepsilon', \beta, \bar\varepsilon'')$ the length of which would be less than

$$L(C_\varepsilon) = s_{\varepsilon'\varepsilon''} + s_{\varepsilon'\beta\varepsilon''} .$$

This, however, would contradict our assumptions on ε' and ε''. Hence C_ε is a basic cycle.

To prove that C_ε can be chosen for D, it remains to be shown that $A(\alpha, C_\varepsilon) = \beta$, i.e. that $s_{\alpha\gamma} < s_{\alpha\beta}$ for every $\gamma \in C_\varepsilon$, $\gamma \neq \beta$. This is self-evident if $\gamma \in v'_\varepsilon \cup v''_\varepsilon$, since $v'_\varepsilon \cup v''_\varepsilon \subset (\alpha, \beta)$. If $\gamma \in v_\varepsilon$, then we have

$$\begin{aligned} s_{\alpha\gamma} &\leq \tfrac{1}{2} (s_{\alpha\varepsilon'} + s_{\alpha\varepsilon''} + s_{\varepsilon'\varepsilon''}) \\ &< \tfrac{1}{2} (s_{\alpha\varepsilon'} + s_{\alpha\varepsilon''} + s_{\varepsilon'\beta\varepsilon''}) = s_{\alpha\beta} , \end{aligned}$$

since $s_{\varepsilon'\varepsilon''} < s_{\varepsilon'\beta\varepsilon''}$.

We have hitherto supposed that $\beta \in T_1(\alpha)$ and verified the theorem in this case. The case $\beta \in T_2(\alpha)$ can be reduced to the previous one, since, by Theorem 2.3, there exists a junction point $\gamma \in (\alpha, \beta)$ for which $\beta \in T_1(\gamma)$.

Remark. If the order $n_{\beta\alpha} - 1$ of the reflection point β is higher than 1, one can see, by inspecting the foregoing proof, that there exist at least $\binom{n_{\beta\alpha}}{2} > n_{\beta\alpha} - 1$ different cycles like C and hence it can be concluded that there are at least $\binom{n_{\beta\alpha}}{2}$ basic cycles to which the theorem applies.

The following corollaries are immediate consequences of Theorem 3.2.

Corollary 3.2.1. If α and β are reflection points of each other, then there exist basic cycles D' and D'' such that $\alpha \in A(D'', D')$, $\beta \in A(D', D'')$.

Let r be the number of basic cycles in A and let the basic cycles be denoted by $D_1, D_2, \ldots, D_r$.

Corollary 3.2.2. The cyclomatic number of A is at most equal to the number of basic cycles, i.e., $m \leq r$. If A is singular, then $m < r$.

This is true for nonsingular networks, according to Theorem 2.7. From the remark made after the proof of Theorem 3.2 and from (2.6) it follows that $m < r$ for singular networks.

3.3. Törnqvist networks

Networks for which $r = m$ will be called *Törnqvist networks*. By Corollary 3.2.2, all Törnqvist networks are nonsingular. In practice, planar networks are usually Törnqvist networks since, as remarked in 1.3 (Euler's formula), the cyclomatic number m equals the number of the faces of the planar network, and the faces are usually the only basic cycles.

A simple example of a network which is not a Törnqvist network is that formed by the edges of a cube. Here $r = 6$, but

$$m = 12 - 8 + 1 = 5 .$$

By Theorem 3.2 and Corollary 3.2.2, we have the following corollary on Törnqvist networks.

Corollary 3.2.3. In a Törnqvist network every point has precisely one extreme point on each basic cycle. The extreme points α_i of α on the basic cycles D_i, $i = 1, 2, \ldots, m$ are all different and identical with the reflection points of α.

We shall make use of the shorter notation

$$A_{ij} = A(D_i, D_j) , \quad i, j = 1, 2, \ldots, r$$

for the set of extreme points of D_i on D_j. Hence, by the definition of a basic cycle, we have

$$A_{ii} = D_i\,, \qquad i = 1, 2, \ldots, r\,.$$

Correspondingly we denote

$$s_{\alpha i} = s(\alpha\,, D_i)\,, \qquad i = 1, 2, \ldots, r\,,$$

$$s_{ij} = s(D_i\,, D_j)\,, \qquad i\,, j = 1, 2, \ldots, r\,,$$

$$a_{ij} = L(A_{ij})\,, \qquad i\,, j = 1, 2, \ldots, r\,.$$

Let us study the structure of the sets A_{ij} more closely. First consider a path v which does not contain junction points, or reflection points of junction points, as inner points; in particular, we shall examine the extreme points of v on a basic cycle D.

Let $\alpha \in v$ have as an extreme point on D a point β which, according to Theorem 3.1, is also a reflection point of α. If $\beta \in T_2(\alpha)$ for all $\alpha \in v$, then $A(v, D)$ consists of a finite number of distinct points on D all of which are reflection points of junction points. This follows from Theorem 2.3.

On the other hand, if for some $\alpha \in v$ we have $\beta \in T_1(\alpha)$, it is seen directly from Theorem 2.4 that $A(D, v) = v$ and $A(v, D) \subset T_1(v)$. The set $A(v, D)$ then consists of one or more paths each of which has the same length $L(v)$, according to Corollary 2.5.1. In particular, if A is a Törnqvist network, then, according to Corollary 3.2.3, $A(v, D)$ consists of a single component, and $L(A(D, v)) = L(A(v, D)) = L(v)$.

Since we may now suppose that D_i consists of paths $v_h, h = 1, 2, \ldots, p$ none of which contains junction points, or reflection points of junction points, as inner points, it follows that

$$s_{ij} = \max_h s(v_h\,, D_j)\,;$$

hence we have arrived at the following theorem:

THEOREM 3.3. The set A_{ij} consists of a finite number of paths and distinct points on D_j. If A is a Törnqvist network, then $a_{ji} = a_{ij}$.

Usually A_{ij} is a proper subset of D_j. There are, however, cases in which $A_{ij} = D_j$. For instance, in the network formed by the edges of

a cube we actually have the situation $A_{ij} = D_j$ for the opposite faces of the cube. This case is never possible in Törnqvist networks, as stated in the following theorem:

THEOREM 3.4. In a Törnqvist network A the set A_{ij} is always a proper subset of D_j when $i \neq j$.

PROOF. Let us suppose, to the contrary, that the Törnqvist network A contains two basic cycles D_i, D_j such that $A_{ij} = D_j$; it will be shown that this leads to a contradiction. It is first verified that for all points α, β on D_j we have $s_{\alpha\beta} = s_{\alpha_i\beta_i}$ where α_i and β_i are the extreme points of α and β on D_i. It may be noted that α_i and β_i are unique, by Corollary 3.2.3.

Next, let E be the set of all special points in A, and consider the set $E_j = D_j \cap (E \cup T(E))$. This set consists of a finite number, say p, of points. We denote these points by $\gamma_1, \gamma_2, \ldots, \gamma_p$. It is assumed that γ_h and γ_{h+1} $(h = 1, 2, \ldots, p,\ \gamma_{p+1} = \gamma_1)$ are consecutive points on D_j.

Let α be an inner point of the path $(\gamma_h, \gamma_{h+1}) \cap D_j$ and let α_i be the extreme point of α on D_i. Conversely, since by the antithesis $A_{ij} = D_j$, and therefore $s_{\alpha\alpha_i} = s_{ij}$, α is seen to be the extreme point of α_i on D_j and thus $\alpha \in T(\alpha_i)$, by Theorem 3.1. From this we deduce that $\alpha \in T_1(\alpha_i)$, since if $\alpha \in T_2(\alpha_i)$, we would, by Theorem 2.3. have $\alpha \in T(E)$, which is impossible according to the definition of E_j. Thus $\alpha \in T_1(\alpha_i)$ and from Theorem 2.4, 3° it follows that $s_{\alpha\beta} = s_{\alpha_i\beta_i}$ when both of the points α, β are on the same path $(\gamma_h, \gamma_{h+1}) \cap D_j$. Let the extreme point of γ_h on D_i be γ_{hi}. Then also $s_{\gamma_h\gamma_{h+1}} = s_{\gamma_{hi}\gamma_{h+1,i}}$.

The paths $(\gamma_{hi}, \gamma_{h+1,i}) \cap D_i$, $h = 1, 2, \ldots, p$ can meet only in the points $\gamma_{hi}, h = 1, 2, \ldots, p$, since if $\delta \neq \gamma_{hi}, h = 1, 2, \ldots, p$ were a common point of two or more such paths, then the extreme point δ_j of δ on D_j would not be unique, and this would contradict Corollary 3.2.3. We thus conclude that the points $\gamma_{1i}, \gamma_{2i}, \ldots, \gamma_{pi}$ follow each other in this order along D_i, and that

$$A_{ji} = D_i \cap \bigcup_{h=1}^{p} (\gamma_{hi}\,, \gamma_{h+1,i}) = D_i\,.$$

Since $s_{\gamma_h\gamma_{h+1}} = s_{\gamma_{hi}\gamma_{h+1,i}}$, $h = 1, 2, \ldots, p$, we have $s_{\gamma_h\gamma_g} = s_{\gamma_{hi}\gamma_{gi}}$, $h, g = 1, 2, \ldots, p$ and finally $s_{\alpha\beta} = s_{\alpha_i\beta_i}$, $\alpha \in D_j$, $\beta \in D_j$, since D_i is a basic cycle.

We shall now prove that D_i and D_j cannot have any points in common. Indeed, if α were such a common point and α_i the extreme point of α on D_i, it would follow that $s_{\alpha\alpha_i} = s_{\alpha i} = s_{ii} = s_{ij}$, since $\alpha \in A_{ij}$. This would mean that α is an extreme point of α_i both on D_i and D_j, which again contradicts Corollary 3.2.3.

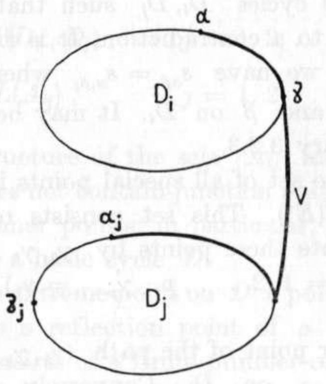


Fig. 3.3

Let us consider an arbitrary point $\alpha \in D_i$ and its extreme point α_j on D_j. Let v be an (α, α_j)-path. Since D_i and D_j cannot have any points in common, there exists a junction point $\gamma \in D_i$, in which v takes off from D_i. The path v then has no points except γ in common with D_i between γ and α_j. Since it has been shown that $A_{ji} = D_i$, i.e., $s_{\alpha\alpha_j} = s_{\gamma\gamma_j} = s_{ij}$ and $s_{\alpha\gamma} = s_{\alpha_j\gamma_j}$, there must exist a (γ, γ_j)-path v' passing through α_j. It follows that v and v' coincide between γ and α_j. Hence v' and D_i have no common points except the junction point γ. Because γ is the extreme point of γ_j on D_i, we must have $n_{\gamma\gamma_j} \geq 3$. However, since a Törnqvist network is always nonsingular, this is impossible. Thus the assumption $A_{ij} = D_j$ is false.

4. THE DISTANCE DISTRIBUTION

4.1. Formulation of the problem

The distance distribution in a network can be investigated starting from different basic assumptions. One natural starting point is as follows.

In a network A we choose two points α and β at random independently of each other. The expression 'at random' means that the probability of choosing α (and likewise β) from an arbitrary subnetwork B of A equals $a^{-1}L(B)$. In other words, we assume the random points α and β to be uniformly and independently distributed over A. The distance $s_{\alpha\beta}$ between α and β will then be a random variable having a certain distribution depending only on A. Our objective is to determine this distribution.

We shall use the notations $f(s)$ and $F(s)$ for the density and distribution function, respectively, of the random variable $s_{\alpha\beta}$. Similarly, $f_\alpha(s)$ and $F_\alpha(s)$ will denote the conditional density and distribution functions, given α.

4.2. The functions $L_\alpha(s)$, $F_\alpha(s)$, $f_\alpha(s)$

The points β whose distance from a fixed point α is less than s constitute a subnetwork of A the length of which will be denoted by $L_\alpha(s)$, i.e.,

$$L_\alpha(s) = L\{\beta \mid s_{\alpha\beta} \leq s\} .$$

Obviously, the conditional distribution function $F_\alpha(s)$ is

$$F_\alpha(s) = P\{s_{\alpha\beta} \leq s \mid \alpha\} = a^{-1}L_\alpha(s) .$$

Let us give to s a positive increment Δs and form the corresponding increment of $L_\alpha(s)$. By the definition of $n_\alpha(s)$ we have, for a sufficiently small Δs,

$$L_\alpha(s + \Delta s) - L_\alpha(s) = L\{\beta \mid s < s_{\alpha\beta} \leq s + \Delta s\} = n_\alpha(s)\,\Delta s .$$

Hence the function $L_\alpha(s)$ is continuous on the right and possesses a right hand derivative $n_\alpha(s)$ at every point s.

In a similar manner it can be shown that $L_\alpha(s)$ is also continuous on the left and has the left hand derivative $n_\alpha(s)$ at every point s except at some of the points (2.3).

Since $F_\alpha(s)$ is up to a constant factor a^{-1} identical with $L_\alpha(s)$, it follows that $F_\alpha(s)$ is continuous at every point s and has the derivative $a^{-1}n_\alpha(s)$ for every value of s except for some of the values (2.3). The conditional density function $f_\alpha(s)$ can therefore be written in the form

$$f_\alpha(s) = a^{-1}n_\alpha(s)\,. \tag{4.1}$$

4.3. The density function $f(s)$

The distribution function $F(s)$ of the random variable $s_{\alpha\beta}$ can be written as the expected value

$$F(s) = \underset{\alpha}{E}F_\alpha(s)\,.$$

The corresponding density function $f(s)$ is then by (4.1)

$$f(s) = \underset{\alpha}{E}f_\alpha(s) = a^{-1}\underset{\alpha}{E}n_\alpha(s)\,.$$

Observe that differentiation under $\underset{\alpha}{E}$ is permitted in this case (cf. Cramér [2], p. 67). Using the expression (2.5) for $n_\alpha(s)$ we find

$$af(s) = 2\,\varphi(s) + \sum_{k=1}^{m_0}(n_{\varepsilon_k} - 2)\,\underset{\alpha}{E}\varphi(s - s_{\alpha\varepsilon_k})$$
$$- 2\,\underset{\alpha}{E}\sum_{\beta\in T(\alpha)}(n_{\beta\alpha} - 1)\,\varphi(s - s_{\alpha\beta})\,.$$

Observing that, for any fixed β and random α,

$$F_\beta(s) = P\{s_{\alpha\beta} \le s\} = \underset{\alpha}{E}\varphi(s - s_{\alpha\beta})$$

we finally obtain

$$af(s) = 2\,\varphi(s) + \sum_{k=1}^{m_0}(n_{\varepsilon_k} - 2)\,F_{\varepsilon_k}(s) - 2\,e(s) \tag{4.2}$$

where

$$e(s) = \underset{\alpha}{E}\sum_{\beta\in T(\alpha)}(n_{\beta\alpha} - 1)\,\varphi(s - s_{\alpha\beta})\,. \tag{4.2)'$$

To make the expression of $f(s)$ more explicit we need suitable formulas for $F_\alpha(s)$ and $e(s)$. $F_\alpha(s)$ can be evaluated by integrating both sides of (4.1) and using again (2.5):

$$aF_\alpha(s) = a\int_0^s f_\alpha(s)\,ds = \int_0^s n_\alpha(s)\,ds \tag{4.3}$$
$$= 2\,s\varphi(s) + \sum_{k=1}^{m_0}(n_{\varepsilon_k} - 2)\,(s - s_{\alpha\varepsilon_k})\,\varphi(s - s_{\alpha\varepsilon_k})$$
$$- 2\sum_{\beta\in T(\alpha)}(n_{\beta\alpha} - 1)\,(s - s_{\alpha\beta})\,\varphi(s - s_{\alpha\beta})\,.$$

The construction of a suitable expression for $e(s)$ is more difficult. From (4.2)′ and (2.6) it can be seen that $e(s)$ is a monotonically increasing and non-negative function of s and that always $e(s) \le m$.

In particular, if A is a tree then $e(s) = 0$ for all s and we have simply

$$af(s) = 2\,\varphi(s) + \sum_{k=1}^{m_0}(n_{\varepsilon_k} - 2)\,F_{\varepsilon_k}(s)$$

where $F_{\varepsilon_k}(s)$ is now

$$F_{\varepsilon_k}(s) = 2\,a^{-1}s\varphi(s) + a^{-1}\sum_{h=1}^{m_0}(n_{\varepsilon_h} - 2)\,(s - s_{\varepsilon_k\varepsilon_h})\,\varphi(s - s_{\varepsilon_k\varepsilon_h})\,.$$

4.4. Decomposition of $e(s)$

We shall consider $e(s)$ more closely only for Törnqvist networks. In the rest of this chapter it is thus supposed that the network A is a Törnqvist network, i.e., the cyclomatic number m equals the number r of basic cycles. In Chapter 5 we shall return to the general case.

Due to the nonsingularity of all Törnqvist networks, we always have $n_{\beta\alpha} \le 2$, and (4.2)′ can be written as

$$e(s) = \underset{\alpha}{E}\sum_{\beta\in T(\alpha)}\varphi(s - s_{\alpha\beta})\,.$$

By Corollary 3.2.3, we may simply sum over the basic cycles, thus obtaining

$$e(s) = \underset{\alpha}{E}\sum_{i=1}^{m}\varphi(s - s_{\alpha i}) = \sum_{i=1}^{m}\underset{\alpha}{E}\varphi(s - s_{\alpha i})$$

where $s_{\alpha i}$ (cf. 3.3) is the distance from α to its extreme point α_i on D_i. If we denote

$$e_i(s) = \underset{\alpha}{E}\varphi(s - s_{\alpha i}),$$

it can be seen that $e(s)$ has been decomposed into a sum of m terms:

$$e(s) = \sum_{i=1}^{m} e_i(s). \tag{4.4}$$

We then consider a particular $e_i(s)$ and write it into the form

$$e_i(s) = a^{-1}L_i(s) \tag{4.5}$$

where

$$L_i(s) = a\underset{\alpha}{E}\varphi(s - s_{\alpha i}).$$

Denote by $A_i(s)$ the set of all points α in A for which $s_{\alpha i} \leq s$. It follows that

$$\varphi(s - s_{\alpha i}) = \begin{cases} 1 & \text{if } \alpha \in A_i(s) \\ 0 & \text{if } \alpha \notin A_i(s) \end{cases}$$

from which we deduce that

$$L_i(s) = L\{A_i(s)\}.$$

Hence $L_i(s)$ is a function associated with the basic cycle D_i in the same way that the function $L_\alpha(s)$ was associated with the point α.

4.5. The function $n_i(s)$

With every basic cycle D_i we shall associate a function $n_i(s)$ analogous to the functions $n_\alpha(s)$ that we have associated with the single points α. If s has none of the values

$$\begin{aligned} s &= s_{ij}\,, \quad j = 1, 2, \ldots, m \\ s &= s_{\varepsilon_k i}\,, \quad k = 1, 2, \ldots, m_0 \end{aligned} \tag{4.6}$$

we define $n_i(s)$ as the number of points α for which $s_{\alpha i} = s$. For the values (4.6) we define $n_i(s)$ as the limit on the right of the function previously defined. This is possible, since $n_i(s)$ is a step function with possible discontinuities at the points (4.6).

We shall show that $n_i(s)$ cannot have other discontinuities. Consider the jump of $n_i(s)$ caused by a point α for which $s_{\alpha i}$ is not equal to any of the values s_{ij}, $j = 1, 2, \ldots, m$. Then $n_{\alpha\alpha_i}$ (i.e., the relative degree of α with respect to its extreme point α_i on D_i) equals 1. Indeed, if we had $n_{\alpha\alpha_i} > 1$, then α would be an extreme point of D_i on some basic cycle D_j, according to Corollaries 3.2.1 and 3.2.3. In this case $s_{\alpha i} = s_{ij}$, against our assumption about α.

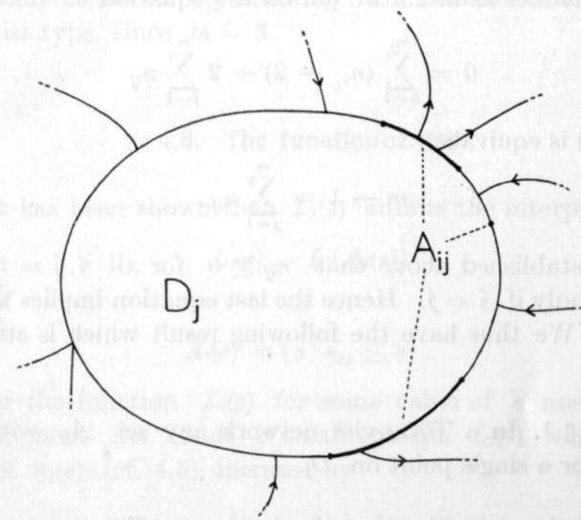
Hence we conclude that the jump of $n_i(s)$ at $s = s_{\alpha i}$ caused by α is

$$(n_\alpha - n_{\alpha\alpha_i}) - n_{\alpha\alpha_i} = n_\alpha - 2\,.$$

In particular, if $s_{\alpha i}$ is not equal to any of the values (4.6), the jump equals zero, since α is then a line point.

It has been verified that the only possible discontinuities of $n_i(s)$ are at the points (4.6). On the other hand, since $n_i(s)$ is, by definition, integer-valued at all points except those of (4.6), and changes its value in these points only, it is continuous and constant between any two consecutive points (4.6). Thus, $n_i(s)$ is a step function similar to $n_\alpha(s)$.

We now consider the jump of $n_i(s)$ at $s = s_{ij}$ caused by the set A_{ij}. We recall that A_{ij} is the set of points on D_j possessing the extreme distance s_{ij} to D_i.



$n_{ij} = 3 \qquad \text{jump} = -2 \times 3 + 3 = -3\,.$

Fig. 4.1

Let us denote the complement set of A_{ij} with respect to D_j by $\bar{A}_{ij}$. According to Theorem 3.3, the set $\bar{A}_{ij}$ consists of a finite number of open paths on D_j. Let the number of connected components in $\bar{A}_{ij}$ be n_{ij}.

Since $A_{ii} = D_i$, we have $\bar{A}_{ii} = \emptyset$ and $n_{ii} = 0$, $i = 1, 2, \ldots, m$. If $i \neq j$, then $n_{ij} > 0$, since A_{ij} is in this case a proper subset of D_j, by Theorem 3.4. Hence, when $i \neq j$, n_{ij} also indicates the number of components in A_{ij}.

Since for every point $\bar{\alpha} \in \bar{A}_{ij}$ we have $s_{\bar{\alpha}i} < s_{ij}$, the jump of $n_i(s)$ at $s = s_{ij}$ caused by A_{ij} is $-2\, n_{ij}$ plus the jumps caused by those junction points of A which belong to A_{ij}. On account of the nonsingularity of A the jump caused by such a junction point α is $n_\alpha - n_{\alpha\alpha_i} = n_\alpha - 2$, and thus of the same form as the jumps caused by the other junction points, as stated above.

As a result of these considerations we obtain for $n_i(s)$ the expression

$$n_i(s) = \sum_{k=1}^{m_0} (n_{\varepsilon_k} - 2)\, \varphi(s - s_{\varepsilon_k i}) - 2 \sum_{j=1}^{m} n_{ij} \varphi(s - s_{ij}) \tag{4.7}$$

for all values of s.

By its definition, $n_i(s)$ vanishes when s becomes large enough. In a similar manner as in 2.8 we obtain the equation

$$0 = \sum_{k=1}^{m_0} (n_{\varepsilon_k} - 2) - 2 \sum_{j=1}^{m} n_{ij}$$

which by (2.2) is equivalent to

$$m - 1 = \sum_{j=1}^{m} n_{ij} \, .$$

It has been established above that $n_{ij} \geq 0$ for all $i, j = 1, 2, \ldots, m$, and $n_{ij} = 0$ only if $i = j$. Hence the last equation implies that $n_{ij} = 1$ when $i \neq j$. We thus have the following result which is stronger than Theorem 3.4.

THEOREM 4.1. In a Törnqvist network any set A_{ij} with $i \neq j$ is either a path or a single point on D_j.

It is interesting to notice that if A is not a Törnqvist network, the sets A_{ij} actually may possess several components. For instance, in Fig. 4.2, the network consists of four basic cycles $D_1 = (\alpha', \alpha'', \beta'', \beta')$,

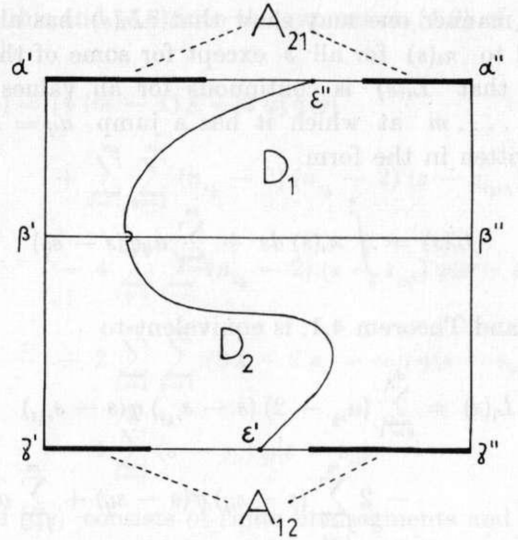


Fig. 4.2

$D_2 = (\beta', \beta'', \gamma'', \gamma')$, $D_3 = (\alpha', \varepsilon'', \varepsilon', \gamma')$, $D_4 = (\varepsilon'', \alpha'', \gamma'', \varepsilon')$, and we have two components in A_{12} and A_{21}. The network is not, however, of the Törnqvist type, since $m = 3$.

4.6. The function $L_i(s)$

In 4.4 it has been shown that $L_i(s)$ admits the interpretation

$$L_i(s) = L\{A_i(s)\}$$

where

$$A_i(s) = \{\alpha \mid s_{\alpha i} \leq s\} \, .$$

We consider the function $L_i(s)$ for some value of s and give to s a positive increment Δs. If Δs is small enough, $L_i(s)$ will then, by the definition of $n_i(s)$ (cf. 4.5), increase by

$$L_i(s + \Delta s) - L_i(s) = L\{\alpha \mid s < s_{\alpha i} \leq s + \Delta s\} = n_i(s)\, \Delta s \, .$$

Hence $L_i(s)$ has $n_i(s)$ for derivative on the right for all s.

In a similar manner one may show that $L_i(s)$ has also a left hand derivative equal to $n_i(s)$ for all s except for some of the values (4.6).

We observe that $L_i(s)$ is continuous for all values of s except $s = s_{ij}, j = 1, 2, \ldots, m$ at which it has a jump $a_{ij} = L(A_{ij})$. Hence $L_i(s)$ can be written in the form

$$L_i(s) = \int_0^s n_i(s)\, ds - \sum_{j=1}^{m} a_{ij}\varphi(s - s_{ij})$$

which, by (4.7) and Theorem 4.1, is equivalent to

$$L_i(s) = \sum_{k=1}^{m_0} (n_{\varepsilon_k} - 2)(s - s_{\varepsilon_k i})\,\varphi(s - s_{\varepsilon_k i}) - 2\sum_{\substack{j=1 \\ j\neq i}}^{m} (s - s_{ij})\,\varphi(s - s_{ij}) + \sum_{j=1}^{m} a_{ij}\varphi(s - s_{ij})\,. \tag{4.8}$$

4.7. The density function $f(s)$ in a Törnqvist network

In order to find an explicit expression for the density function $f(s)$ of $s_{\alpha\beta}$, we have only to combine the results of the previous sections. According to (4.4) and (4.5), we have

$$e(s) = a^{-1}\sum_{i=1}^{m} L_i(s)\,.$$

Using the general formula (4.2) of $f(s)$, we then obtain

$$af(s) = 2\,\varphi(s) + \sum_{k=1}^{m_0} (n_{\varepsilon_k} - 2)\,F_{\varepsilon_k}(s) - 2\,a^{-1}\sum_{i=1}^{m} L_i(s)\,. \tag{4.9}$$

The factor $F_{\varepsilon_k}(s)$ is obtained from (4.3) on taking $\alpha = \varepsilon_k$ and noting that the nonsingularity of the network implies $n_{\beta\alpha} = 2$. Thus

$$aF_{\varepsilon_k}(s) = 2\,s\varphi(s) + \sum_{h=1}^{m_0} (n_{\varepsilon_h} - 2)(s - s_{\varepsilon_k\varepsilon_h})\,\varphi(s - s_{\varepsilon_k\varepsilon_h}) - 2\sum_{i=1}^{m} (s - s_{\varepsilon_k i})\,\varphi(s - s_{\varepsilon_k i})\,.$$

By substituting this and (4.8) into the expression (4.9) of $f(s)$ we obtain the final formula

$$\begin{aligned} a^2 f(s) = {} & [4\,(m-1)\,s + 2\,a]\,\varphi(s) \\ & + \sum_{k=1}^{m_0}\sum_{h=1}^{m_0} (n_{\varepsilon_k} - 2)(n_{\varepsilon_h} - 2)(s - s_{\varepsilon_k\varepsilon_h})\,\varphi(s - s_{\varepsilon_k\varepsilon_h}) \\ & - 4\sum_{i=1}^{m}\sum_{k=1}^{m_0} (n_{\varepsilon_k} - 2)(s - s_{\varepsilon_k i})\,\varphi(s - s_{\varepsilon_k i}) \\ & + 2\sum_{i=1}^{m}\sum_{j=1}^{m} (2\,s - 2\,s_{ij} - a_{ij})\,\varphi(s - s_{ij}) \\ & - 4\sum_{i=1}^{m} (s - s_{ii})\,\varphi(s - s_{ii})\,. \end{aligned} \tag{4.10}$$

The graph of $f(s)$ consists of finite line segments and it has at most $\binom{m+1}{2} + 1$ discontinuities, located (if present) at the points $s = 0$ and $s = s_{ij},\ i, j = 1, 2, \ldots, m$.

4.8. Examples

We shall apply the previous results to two elementary cases.

1° We first consider a network consisting of a single edge v. Let the end points of v be ε_1 and ε_2. Since v is a tree, $e(s)$ vanishes identically and we have

$$af(s) = 2\,\varphi(s) - \sum_{k=1}^{2} F_{\varepsilon_k}(s)$$

where

$$aF_{\varepsilon_k}(s) = 2\,s\varphi(s) - s\varphi(s) - (s - a)\,\varphi(s - a)\,, \quad k = 1, 2.$$

Hence

$$f(s) = 2\,a^{-2}(a - s)\,[\varphi(s) - \varphi(s - a)]\,.$$

It can be seen that $f(s)$ decreases linearly from $2\,a^{-1}$ to 0 in the interval $(0, a)$. Elsewhere $f(s)$ vanishes. The expected value of the distance between two random points is $\dfrac{a}{3}$.

2° As a second example let us consider a single cycle $A = D_1$. In this case

$$af(s) = 2\,\varphi(s) - 2\,a^{-1}L_1(s)$$

where

$$L_1(s) = a\varphi\left(s - \frac{a}{2}\right),$$

and finally

$$f(s) = 2\,a^{-1}\left[\varphi(s) - \varphi\left(s - \frac{a}{2}\right)\right].$$

The distance is uniformly distributed over the interval $\left(0\,,\,\frac{a}{2}\right)$.

5. GENERALIZATION AND PRACTICAL SOLUTION

5.1. Generalization of the problem

In Chapter 4 we worked out an expression for the distribution of the distance between two independently chosen random points in a Törnqvist network. The uniform distribution of these points was a basic assumption, as stated in 4.1.

The solution presented can be applied in practice without any difficulty. It is, however, a drawback that the solution is applicable only to networks of a certain type, namely, to Törnqvist networks. Also, the assumption about the uniform distribution of the random points over the whole network is in most cases unrealistic.

For these reasons we have developed another method, based on more general assumptions and suitable for all networks. The general method cannot, of course, be as economic in computations as that presented in Chapter 4, but its larger applicability can be considered a decisive advantage.

The generalized formulation of the problem is as follows.

Our aim is still to study the distance $s_{\alpha\beta}$ between two independently chosen random points α, β in a network A. However, α will now be interpreted as a *source* sending traffic with a certain *intensity*, whereas β is interpreted as a *sink* similarly receiving traffic. In a precise form our assumptions are:

The network A is partitioned into subnetworks $A_1, A_2, \ldots, A_k$. (In the partition the points of division are considered as belonging to all adjacent subnetworks.) The probability of choosing the source α from within A_i is

$$P\{\alpha \in A_i\} = p_i\,, \quad i = 1, 2, \ldots, k$$

and the probability of choosing the sink β from within A_j is

$$P\{\beta \in A_j\} = q_j\,, \quad j = 1, 2, \ldots, k\,.$$

The choices of the source and the sink are independent, i.e.,

$$P\{\alpha \in A_i\,, \beta \in A_j\} = p_i\,q_j\,, \quad i\,, j = 1, 2, \ldots, k\,.$$

Inside the subnetworks the distribution of the random points is uniform.

The above assumptions imply that for each subnetwork B_i of A_i, and each subnetwork B_j of A_j, we have

$$P\{\alpha \in B_i\} = p_i\,\frac{L(B_i)}{L(A_i)}\,, \quad P\{\beta \in B_j\} = q_j\,\frac{L(B_j)}{L(A_j)}\,,$$

$$P\{\alpha \in B_i\,, \beta \in B_j\} = p_i\,q_j\,\frac{L(B_i)L(B_j)}{L(A_i)L(A_j)}\,.$$

The generalized distribution function $\bar{F}(s)$ of $s_{\alpha\beta}$ is then

$$\bar{F}(s) = P\{s_{\alpha\beta} \le s\} = \sum_{i=1}^{k}\sum_{j=1}^{k} p_i\,q_j\,P\{s_{\alpha\beta} \le s \mid \alpha \in A_i\,, \beta \in A_j\}$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{k} p_i\,q_j\,\bar{F}_{ij}(s)$$

where $\bar{F}_{ij}(s)$ is the conditional distribution function

$$\bar{F}_{ij}(s) = P\{s_{\alpha\beta} \le s \mid \alpha \in A_i\,, \beta \in A_j\}\,.$$

The corresponding density function $\bar{f}(s)$ takes the form

$$\bar{f}(s) = \sum_{i=1}^{k}\sum_{j=1}^{k} p_i\,q_j\,\bar{f}_{ij}(s)$$

where $\bar{f}_{ij}(s)$ is the density function corresponding to the conditional distribution function $\bar{F}_{ij}(s)$.

The intensities t_i, u_j, describing the sending and receiving of traffic in the subnetworks $A_1, A_2, \ldots, A_k$, are (as a consequence of the uniform distribution within subnetworks)

$$t_i = \frac{p_i}{L(A_i)}, \quad i = 1, 2, \ldots, k$$

$$u_j = \frac{q_j}{L(A_j)}, \quad j = 1, 2, \ldots, k .$$

Hence the intensities have to satisfy the conditions

$$\sum_{i=1}^{k} t_i L(A_i) = \sum_{j=1}^{k} u_j L(A_j) = 1 .$$

Using the intensities we may write $\bar{f}(s)$ in the form

$$\text{(5.1)} \qquad \bar{f}(s) = \sum_{i=1}^{k} \sum_{j=1}^{k} t_i u_j L(A_i) L(A_j) \bar{f}_{ij}(s) = \sum_{i=1}^{k} t_i L(A_i) \bar{f}_i(s)$$

where $\bar{f}_i(s)$ is the density function

$$\bar{f}_i(s) = \sum_{j=1}^{k} q_j \bar{f}_{ij}(s) = \sum_{j=1}^{k} u_j L(A_j) \bar{f}_{ij}(s)$$

corresponding to the conditional distribution function

$$\bar{F}_i(s) = P\{s_{\alpha\beta} \leq s \mid \alpha \in A_i\} .$$

Hence the density function $\bar{f}(s)$ can be computed as a weighted sum of the conditional densities $\bar{f}_{ij}(s)$. The functions $\bar{f}_{ij}(s)$ can be determined using the results presented in Chapters 2 and 4, since we have assumed the distribution of α and β to be uniform inside the subnetworks A_1, $A_2, \ldots, A_k$.

5.2. Solution of the generalized problem

We shall describe only the main principles of the solution which is based on the assumptions made in 5.1.

We suppose that the intensities are constant on each edge. The reflection points of the junction points divide the edges of A into paths, to which Corollary 2.5.1 applies. We consider these paths, as well as edges without reflection points, as the elements $A_1, A_2, \ldots, A_k$ of a partition of A; they will be called *basic paths*.

According to Corollary 2.5.1, the reflection set $T(A_i)$, $i = 1, 2, \ldots, k$, consists of distinct points and of certain paths (reflection paths)

$$A_{i1}, A_{i2}, \ldots, A_{ik_i}$$

each of which has the length $L(A_i)$. According to (2.6), we have $k_i \leq m$. On each A_{ih}, $h = 1, 2, \ldots, k_i$, the intensities are constant, since otherwise A_{ih} would contain a junction point which would have a reflection point on the basic path A_i.

Fig. 5.1 represents a network A and its partition into the basic paths A_i. The reflection points of the junction points are indicated by stars (∗) and the reflection paths of one A_i by heavier lines.

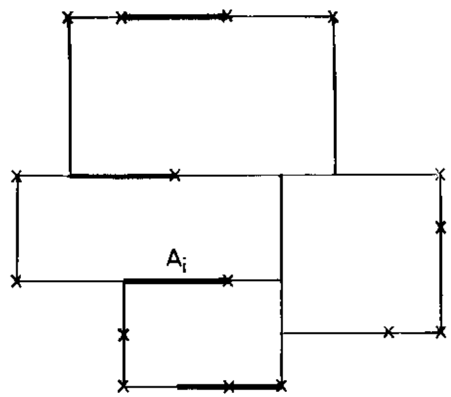


Fig. 5.1

The conditional density function $\bar{f}_i(s)$ can now be computed using a partition $B_1, B_2, \ldots$ the elements of which are

1° A_i,

2° A_{ih}, $h = 1, 2, \ldots, k_i$,

3° the remaining basic paths, and parts of such paths.

When computing $\bar{f}_i(s)$ we thus use a special partition $\{B_j\}$, $j = 1, 2, \ldots$ which, for the most part, coincides with the partition $A_1, A_2, \ldots, A_k$. The conditional densities $\bar{f}_{ij}(s)$, of which $\bar{f}_i(s)$ is a weighted sum, are of three types corresponding to the three types (1°, 2°, 3°, as listed above) of B_j, to which the sink β can belong.

Let us denote $L(A_i) = c_1$, $L(B_j) = c_2$, and let the minimum distance between the points of A_i and B_j be c_3. The density functions $\bar{f}_{ij}(s)$ are then as follows. (Each type can be easily derived using the densities of the examples 1° and 2° in 4.8 as conditional densities.)

1° If $B_j = A_i$, we have

$$\bar{f}_{ij}(s) = 2\,c_1^{-2}(c_1 - s)\,[\varphi(s) - \varphi(s - c_1)]\,.$$

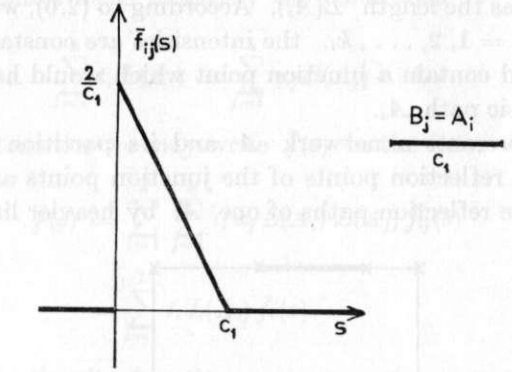


Fig. 5.2

2° If $B_j = A_{ih}$, $h = 1, 2, \ldots, k_i$, in which case $c_2 = c_1$, we have

$$\bar{f}_{ij}(s) = 2\,c_1^{-2}(s - c_3)\,[\varphi(s - c_3) - \varphi(s - c_1 - c_3)]\,.$$

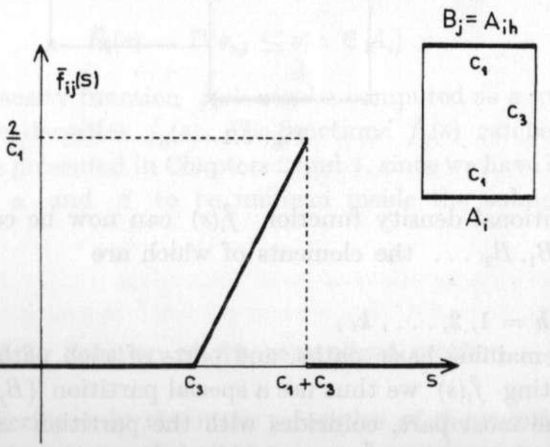


Fig. 5.3

3° If B_j is one of the remaining basic paths, we have

$$\bar{f}_{ij}(s) = c_1^{-1}c_2^{-1}\,[\lambda(s - c_3) - \lambda(s - c_1 - c_3) - \lambda(s - c_2 - c_3) + \lambda(s - c_1 - c_2 - c_3)]$$

where $\lambda(s) = s\varphi(s)$.

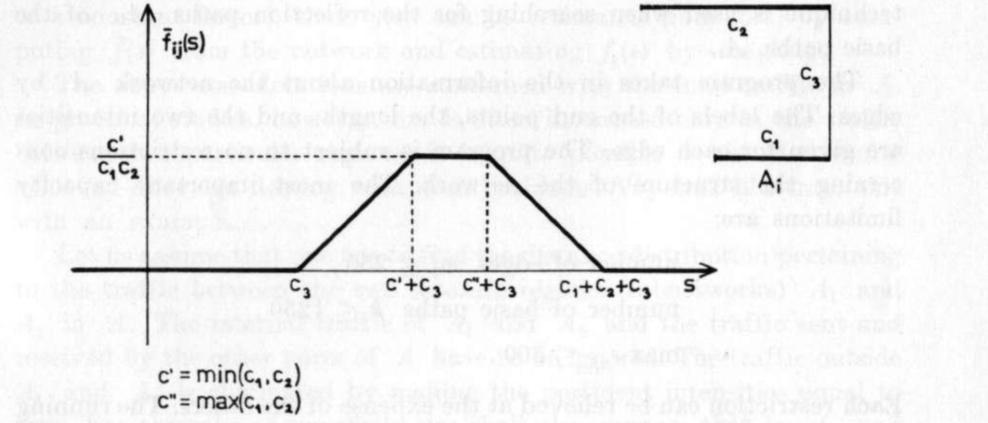


Fig. 5.4

Since $\bar{f}(s)$ is a weighted sum of functions $\bar{f}_{ij}(s)$ of the types described above, we conclude that the graph of $\bar{f}(s)$ is of the same form as that of $f(s)$ in Chapter 4, and it has a finite number of discontinuities.

5.3. Computer program

A general program for the evaluation of the density function $\bar{f}(s)$ of an arbitrary network A has been written for the computer Elliott 803.

The program works according to the principle explained in 5.2. In order to find the basic paths, it first locates all reflection points of junction points. Using the partition thus obtained, it then computes the density function $\bar{f}(s)$. The values obtained are exact when the data concerning A are given as integers.

In the process of locating the reflection points of the junction points a technique of »marching point sets» is employed. It may be described as follows.

From the junction point α a fictitious set of »marching points» is dispatched in all directions with a constant speed. The »marching set» follows all paths of the network and splits into subsets at the junction points. The points where two or more sets meet again correspond to the reflection points of α. In such a point the marching sets become extinct, if the reflection point is not at the same time a junction point. This process simulates the behaviour of the function $n_\alpha(s)$. A similar technique is used when searching for the reflection paths A_{ih} of the basic paths A_i.

The program takes in the information about the network A by edges. The labels of the end points, the length, and the two intensities are given for each edge. The program is subject to no restrictions concerning the structure of the network. The most important capacity limitations are:

$$\text{number of edges } m_1 \leq 250\,,$$
$$\text{number of basic paths } k \leq 1250\,,$$
$$\max s_{\alpha\beta} \leq 500\,.$$

Each restriction can be relieved at the expense of the others. The running time varies considerably depending on the nature and size of the network. The running time for a 5×5 net of squares is 16 minutes, but for a 10×10 net it is nearly 3 hours.

5.4. Applications

We shall close our study with a discussion of the possibilities of applying the previous results.

From the point of view of practical traffic studies, the fact that we have ignored the capacity limitations seems to be a weak point in our model. This is, however, only a seeming deficiency, since traffic congestion can be accounted for by measuring the distances in time.

Another limitation seems to be that the amount of traffic between two regions is assumed to depend only on the sizes and the traffic intensities of these regions. There is no accounting for the distance between the regions. An arbitrary gravitation law depending on the distance can, however, be added to our model by weighting the density function $\bar{f}(s)$ with a given (usually monotonically decreasing) weight function $g(s)$. Hence the density function $f_g(s)$ induced by the gravitation law $g(s)$ is

$$f_g(s) = \frac{\bar{f}(s)\, g(s)}{\int_0^\infty \bar{f}(s)\, g(s)\, ds}\,. \tag{5.2}$$

Conversely, if an observable traffic flow obeys a law of this kind with an unknown gravitation law $g(s)$ (which is obviously determined only up to a constant factor), then $g(s)$ can be estimated from (5.2) by computing $\bar{f}(s)$ from the network and estimating $f_g(s)$ by sampling.

The above considerations are concerned with the total traffic in A. In practical studies, however, one is often interested only in the traffic between two particular regions. It is quite possible to use the results achieved above in dealing with such problems. We shall illustrate this with an example.

Let us assume that one has to find the distance distribution pertaining to the traffic between the two separate regions (subnetworks) A_1 and A_2 in A. The internal traffic of A_1 and A_2 and the traffic sent and received by the other parts of A have to be ignored. The traffic outside A_1 and A_2 is eliminated by making the pertinent intensities equal to zero. For the sake of simplicity, we shall also suppose that in A_1 and A_2 the intensities are constant.

We introduce the notations $L(A_1) = a_1$, $L(A_2) = a_2$, $a_0 = a_1 + a_2$. The distribution function $\bar{F}(s)$ can then be written in the form

$$\begin{aligned}
\bar{F}(s) &= P\{s_{\alpha\beta} \leq s\} \\
&= a_1^2 a_0^{-2} P\{s_{\alpha\beta} \leq s \mid \alpha \in A_1\,, \beta \in A_1\} \\
&\quad + a_2^2 a_0^{-2} P\{s_{\alpha\beta} \leq s \mid \alpha \in A_2\,, \beta \in A_2\} \\
&\quad + 2\, a_1 a_2 a_0^{-2} P\{s_{\alpha\beta} \leq s \mid \alpha \in A_1\,, \beta \in A_2 \text{ or } \alpha \in A_2\,, \beta \in A_1\} \\
&= a_1^2 a_0^{-2} \bar{F}_1(s) + a_2^2 a_0^{-2} \bar{F}_2(s) + 2\, a_1 a_2 a_0^{-2} \bar{F}_{12}(s)
\end{aligned}$$

where $\bar{F}_1(s)$ and $\bar{F}_2(s)$ are the distribution functions related to the internal traffic of A_1 and A_2, while $\bar{F}_{12}(s)$ is the distribution function for the traffic between A_1 and A_2. Applying the same argument to the density functions we obtain

$$\bar{f}_{12}(s) = (2\, a_1 a_2)^{-1} [a_0^2 \bar{f}(s) - a_1^2 \bar{f}_1(s) - a_2^2 \bar{f}_2(s)]$$

from which $\bar{f}_{12}(s)$ can be evaluated, since the densities $\bar{f}(s)$, $\bar{f}_1(s)$, and $\bar{f}_2(s)$ are related to the internal traffic of the networks A, A_1, and A_2.

REFERENCES

[1] BERGE, C.: The Theory of Graphs and its Applications, London 1962.

[2] CRAMÉR, H.: Mathematical Methods of Statistics, Princeton 1946.

[3] REIDEMEISTER, K.: Einführung in die Kombinatorische Topologie, Braunschweig 1951.

[4] TÖRNQVIST, L.: On Distribution Functions for Quantities Related to Networks, Bulletin of the International Statistical Institute, Vol. XXXVIII: Part IV, Tokyo 1961.

Printed April 1964