

Filiation of Manuscripts

Peter Buneman

The recovery of trees from measures of dissimilarity

The problem of inferring an evolutionary tree from a set of measurements is one that crops up in various fields, such as biology, palaeography, and archaeology. For example, amino-acid sequences of the same protein extracted from different organisms can be determined, and one can attempt, from the dissimilarities between these sequences, to construct a phylogenetic tree of these organisms. A similar situation occurs when one has a set of manuscripts all directly or indirectly copied from a common original manuscript. One seeks to reconstruct a family tree or 'stemma' of these documents from the errors that the various scribes made in copying one document from another. A frequent starting point in the solution of such a problem is the measurement of a dissimilarity coefficient (DC) between every pair of objects, and to this end one might, for example, count the number of sites at which two protein sequences differ. Similarly, one might count the number of places at which two manuscripts differ. A DC computed from morphological data for taxonomic purposes could also be used.

The object of this paper is to show that there is a method for inferring a tree from a DC which has properties that may make it rather more attractive than other currently available methods. Sokal and Sneath (1963) and Jardine and Sibson (1971) have given detailed accounts of the measurement of DCs and we shall not discuss this further, except to state that there are circumstances in which the measurement of a DC in the manner outlined above for protein chains or manuscripts can obscure useful information provided by the raw data. Thus the methods discussed here, involving DCs, are not the only ones that might be employed nor are they necessarily the best, but it is felt that they might provide a valuable starting point for closer examination of the data. This may be of special use in the reconstruction of a stemma where, for example, the techniques indicated by Maas (1958) for this purpose are impractical for all but a small number of manuscripts.

If we are given a tree, which may be thought of informally as a collection of nodes and links, the assignment of a length to each link will make each node a certain distance from any other node. This distance is the total length of the path between the two nodes, and we shall call such a measure of distance (which is itself a DC) an *additive tree metric*. A DC is not, in general, an additive tree metric: we shall see later that to be so it must satisfy a specific condition. Our problem then, is to find a transformation from the given DC to an additive tree metric. Cavalli-Sforza and Edwards (1967) and Eck and Dayhoff (1966) among others have described methods for finding the *nearest* additive tree metric to a DC. A dissimilarity or *stress* can be defined between two DCs in a number of ways. Subject to this definition,

the nearest additive tree metric (we shall call it Δ_N) is that which minimizes the stress between it and the given DC. To find Δ_N reliably is usually an impossible task: it could involve a search through all the enormous number of possible configurations for the tree and for each configuration optimizing the link lengths. The methods invented for finding Δ_N operate by making a good guess at an initial configuration and then optimizing for a limited type of perturbation of that configuration. Such methods can still be lengthy, and there is no guarantee that they will not get trapped in a local, but not absolute, minimum for the stress. Another objection to Δ_N , given that it can be found, is that it will usually produce a detailed tree even from a DC which does not at all resemble an additive tree metric. In the absence of any theory as to why the data should give a tree, Δ_N may be misleading.

It will help our purposes to give a slightly unusual definition of a tree; and this is done in the next section. Observe that we are not seeking a tree which directly relates the given objects: we may want to reconstruct 'missing' nodes. Moreover we cannot hope to find a root to the tree on the basis of a DC nor can we necessarily find it even from examination of the raw data. What we shall do, therefore, is to define a tree in terms of its links. We can then show that a dissimilarity coefficient will produce a set of links and an additive tree metric, Δ_d . The relative merits of Δ_N and Δ_d are examined in the last section. The mathematics that follows is all very straightforward, but for any graph theoretical terms whose meaning is not obvious the reader is referred to Harary (1969).

TREES DEFINED BY THEIR LINKS

We shall call the finite set of objects on which the given DC has been measured the *base set*. A link in a tree divides the nodes of the tree into two complementary subsets, and by analogy with this we shall define a *split* of the base set to be a pair of non-empty complementary subsets. Thus if S is the base set, σ is a split if $\sigma = \{S^0, S^1\}$ where $S = S^0 \cup S^1$, and $S^0 \cap S^1$ is empty. If two members A and B of S lie in different members of σ , we shall say that σ separates A from B . Suppose that $\sigma_1 = \{S_1^0, S_1^1\}$ and $\sigma_2 = \{S_2^0, S_2^1\}$ are two splits; they are *compatible* if at least one of the intersections $S_1^0 \cap S_2^0$, $S_1^0 \cap S_2^1$, $S_1^1 \cap S_2^0$, $S_1^1 \cap S_2^1$ is empty. If two such intersections were empty the splits would be identical and, of course, any split is compatible with itself.

A *tree* on S can now be defined as a set of distinct, pairwise compatible splits of S . From now on, unless explicitly stated otherwise, we shall use the word 'tree' in this sense. Suppose that $T = \{\sigma_1, \sigma_2, \dots, \sigma_n\}$ is a tree. A *node* of T is a set $\{S_1^{i_1}, S_2^{i_2}, \dots, S_n^{i_n}\}$ where $S_k^{i_k} \in \sigma_k$, and each of the intersections $S_k^{i_k} \cap S_m^{i_m}$ is non-empty. Nodes exist; we can, for example, choose from each split that member which contains some given member of S . $\mathcal{N}(T)$ will be used to designate the set of nodes of T . Two nodes are *linked* by σ_k if they are, in some order,

$$\text{and } \begin{cases} \{S_1^{i_1}, \dots, S_k^0, \dots, S_n^{i_n}\} \\ \{S_1^{i_1}, \dots, S_k^1, \dots, S_n^{i_n}\} \end{cases} .$$

That is to say that two nodes are linked by a split if they differ just on that split. We can now show that this set of nodes and links constitutes a graph-theoretic tree.

Lemma 1

Let N and M be two nodes of T ; then there is a path from N to M .

This means we can find a sequence of nodes N_1, N_2, \dots, N_p such that $N=N_1$, $M=N_p$ and N_k is linked to N_{k+1} for $1 \leq k < p$. Suppose, for convenience, that N and M differ on $\sigma_1 \dots \sigma_k$ so that they can be expressed:

$$M = \{S_1^{i_1}, \dots, S_k^{i_k}, S_{k+1}^{j_{k+1}}, \dots, S_n^{j_n}\}$$

$$N = \{S_1^{j_1}, \dots, S_k^{i_k}, S_{k+1}^{j_{k+1}}, \dots, S_n^{j_n}\}$$

where $i_1 \neq j_1$, and so on. Of the k members of N which do not agree with the corresponding members of M there must be a minimal one (under set inclusion). Suppose that it is $S_k^{i_k}$. Since it is minimal, its complement $S_k^{j_k}$ must intersect $S_1^{j_1}, \dots, S_{k-1}^{j_{k-1}}$. But $S_k^{j_k}$ also intersects $S_{k+1}^{j_{k+1}}, \dots, S_n^{j_n}$ because M is a node. Therefore

$$N_2 = \{S_1^{j_1}, \dots, S_{k-1}^{j_{k-1}}, S_k^{j_k}, \dots, S_n^{j_n}\}$$

is a node and is linked by σ_k to N . Repeating this process we reduce, at each step, the number of disagreements with M and thus get a path from N to M .

Lemma 2

$$|\mathcal{N}(T)| = |T| + 1$$

The proof is by induction on $|T|$ and the result is trivial when T contains just one split. Let N and M be two nodes linked by some member σ of T . The previous lemma assures us that such nodes exist. There can be only one pair of nodes linked by σ , for were there another such pair we would get a violation of compatibility. Removing σ from T therefore reduces the number of nodes by just one; and this gives the inductive step.

Lemma 3

$\mathcal{N}(T)$, linked by T , is a graph-theoretic tree.

It is connected by lemma 1 and has the right Euler number by lemma 2.

The remaining results in this section are given without proof, which is in all cases easy.

Lemma 4

The (graph theoretic) degree of a node is the number of minimal members it contains. A terminal node has one minimal member.

The *support* of a node is the intersection of its members and a *latent* node is one whose support is empty. A tree is *maximal* if there are no splits, compatible with all its members, which it does not contain.

Lemma 5

A terminal node has non-empty support. A latent node has degree greater than two. The nodes of non-empty support partition the base set.

Lemma 6

$$|\mathcal{N}(T)| \leq 2(|S| - 1). \text{ Equality obtains iff } T \text{ is maximal.}$$

Lemma 7

A tree is maximal iff every terminal node has support containing just one member of S and every other node is latent and of degree three.

It may be of value to interpret briefly these preliminary results in terms of the reconstruction of a stemma. The surviving manuscripts constitute the base set. We can infer the former existence of a missing manuscript only when we have two manuscripts, with some common ancestor, neither of which has been copied from the other. Bearing in mind that we do not

necessarily know the root of the stemma, this corresponds to the statement that a latent node must have degree at least three. Moreover we cannot hope to postulate the former existence of manuscripts which have not been copied from, so that each terminal node of the stemma must have a surviving manuscript in its support. It is also desirable, when the data is not adequate to discriminate them, to allow two or more manuscripts to lie at the same node in the stemma; and so we do not insist that there should be a separate node for each member of the base set. A parallel interpretation holds for the reconstruction of phylogenetic trees. This definition of a tree, in terms of compatible splits, is also of use when constructing a stemma, not from a DC but from the actual variations between texts. In an ideal situation each scribal error would give rise to a split, though in practice the problem is very much complicated by some errors obscuring others.

ADDITIVE TREE METRICS

The given dissimilarity coefficient, d , is a function which associates with each pair of points in the base set S a positive real number. It satisfies

$$\begin{aligned}d(A, A) &= 0 \\d(A, B) &\geq 0 \\d(A, B) &= d(B, A)\end{aligned}$$

for all members of A, B of S . Any split σ of S defines an elementary DC, δ_σ , by

$$\begin{aligned}\delta_\sigma(A, B) &= 1 \text{ if } \sigma \text{ separates } A \text{ from } B \\ \delta_\sigma(A, B) &= 0 \text{ otherwise.}\end{aligned}$$

δ_σ , since it satisfies the triangle inequality, is a pseudometric, and it follows that any positive linear combination $\sum \alpha_\sigma \delta_\sigma$ of these elementary DCs is also a pseudometric. If the splits of such a sum are restricted to being the splits of a tree T , then we can define an *additive tree metric* by:

$$\Delta(A, B) = \sum_{\sigma \in T} \alpha_\sigma \delta_\sigma(A, B) \quad (\alpha_\sigma > 0)$$

[An additive tree metric is strictly speaking only a pseudo-metric for it allows distinct members of S to be zero distance apart.]

An additive tree metric on S will define a metric Δ^* on $\mathcal{N}(T)$ by:

$$\Delta^*(N, M) = \sum_{\sigma \in T} \alpha_\sigma \delta_\sigma^*(N, M)$$

where $\delta_\sigma^*(N, M) = 1$ if σ is a link in the path from N to M ;
= 0 otherwise.

Δ^* and Δ , while they are not the same thing (one is defined on S , the other on $\mathcal{N}(T)$), are such that Δ^* defines Δ for $\Delta(A, B) = \Delta^*(N, M)$ whenever A and B are in the respective supports of N and M . Δ^* corresponds to the informal definition of an additive tree metric given in the introduction but Δ as defined is more convenient for our purposes. We now show how a dissimilarity coefficient naturally gives us an additive tree metric. Suppose that d is a DC on S and $\sigma = \{S^0, S^1\}$ is a split. We define

$$\mu_\sigma = \frac{1}{2} \min (d(A, C) + d(B, D) - d(A, B) - d(C, D))$$

for A, B in S^0 and C, D in S^1 .

Lemma 8

If σ_1 and σ_2 are splits such that $\mu_{\sigma_1} > 0$ and $\mu_{\sigma_2} > 0$ then σ_1 and σ_2 are compatible.

If σ_1 and σ_2 were not compatible then we could choose A, B, C, D so that σ_1 separates A and B from C and D , and so that σ_2 separates A and C from B and D . The quantity $d(A, C) + d(B, D) - d(A, B) - d(C, D)$ would have to be strictly positive since $\mu_{\sigma_1} > 0$, but since $\mu_{\sigma_2} > 0$ it would also have to be strictly negative.

From d we can derive $T_d = \{\sigma : \mu_\sigma > 0\}$ and lemma 8 ensures that T_d is a tree. Consequently we can define an additive tree metric Δ_d by:

$$\Delta_d = \sum_{\sigma \in T_d} \mu_\sigma \delta_\sigma.$$

Lemma 9

$$\Delta_d \leq d.$$

For any pair of members A, B of S there is by lemma 1 a path N_1, N_2, \dots, N_p such that A is in the support of N_1 , B is in the support of N_p , and N_k is linked to N_{k+1} by σ_k for $1 \leq k < p$. For each σ_k we may suppose that $\sigma_k = \{S_k^0, S_k^1\}$ with $A \in S_k^0$ and $B \in S_k^1$. Then we can choose, for each k , $1 < k < p$, a member P_k of S such that P_k lies in $S_k^0 \cap S_{k-1}^1$. We can do this because S_k^0 and S_{k-1}^1 are both members of the node N_k and therefore have a non-empty intersection. Thus σ_k separates A from P_{k+1} for $1 \leq k < p-1$ and it separates B from P_k for $1 < k \leq p-1$; so that

$$\mu_{\sigma_1} \leq \frac{1}{2}(d(A, B) + d(A, P_2) - d(B, P_2)),$$

$$\mu_{\sigma_2} \leq \frac{1}{2}(d(A, P_3) + d(B, P_2) - d(A, P_2) - d(B, P_3)),$$

$$\mu_{\sigma_3} \leq \frac{1}{2}(d(A, P_4) + d(B, P_3) - d(A, P_3) - d(B, P_4)),$$

.

.

$$\mu_{\sigma_{p-1}} \leq \frac{1}{2}(d(A, B) + d(B, P_{d-1}) - d(A, P_{d-1})).$$

Adding these we find that $d(A, B) \geq \sum_{k=1}^{p-1} \mu_{\sigma_k}$. We have considered all those splits in T_d which separate A from B and so the result is proved.

Theorem 1

If T_1 and T_2 are trees and $\sum_{\sigma \in T_1} \alpha_\sigma \delta_\sigma = \sum_{\sigma \in T_2} \beta_\sigma \delta_\sigma$ where $\alpha_\sigma > 0$ and $\beta_\sigma > 0$, then $T_1 = T_2$ and $\alpha_\sigma = \beta_\sigma$.

This means that an additive tree metric specifies a unique tree. Let $d = \sum_{\sigma \in T_1} \alpha_\sigma \delta_\sigma$. If A, B, C, D are chosen so that some split in T_1 separates A and B from C and D then the quantity

$$\frac{1}{2}(d(A, C) + d(B, D) - d(A, B) - d(C, D))$$

is, by the compatibility of splits in T_1 , the sum of the α_σ for exactly those splits which separate A and B from C and D . From this we deduce that for all splits σ of T_1 , $\mu_\sigma \geq \alpha_\sigma$ and hence $T_1 \subseteq T_d$. If the inclusion were strict we would have $\Delta_d > d$ contrary to the previous lemma. If then any of the inequalities $\mu_\sigma > \alpha_\sigma$ were strict we would similarly violate that lemma. We conclude that $T_1 = T_d = T_2$ and $\alpha_\sigma = \mu_\sigma = \beta_\sigma$.

A necessary and sufficient condition that d is an additive tree metric is now given. It is that for all members A, B, C, D of S ,

$$d(A, B) + d(C, D) \leq \max(d(A, C) + d(B, D), d(B, C) + d(A, D))$$

and we shall refer to it as the *four-point condition*. Before proving this assertion we can make some elementary observations about the four-point condition.

If we put $C = D$ in the expression above, it reduces to the triangle inequality, so that a DC which satisfies the four-point condition is necessarily a pseudo-metric. Each sum in this expression corresponds to one of the three distinct ways of partitioning a set of four members into two subsets each of two members. Notice that the four-point condition is equivalent to the condition that two of these sums are equal and not less than the third. It is apparent that there is a resemblance in form between the four-point condition and the ultrametric inequality [$d(A, B) \leq \max(d(A, C), d(B, C))$]; and we note the following result, which can be proved in various ways:

Lemma 10

An ultrametric satisfies the four-point condition.

We now prove our assertion that the four-point condition is necessary and sufficient for a DC to be an additive tree metric.

Theorem 2

$\Delta_d = d$ iff d satisfies the four-point condition.

To establish this we shall first show that Δ_d itself satisfies the four-point condition. The converse is proved by showing that the four-point condition implies that $\Delta_d \geq d$; and this, in view of lemma 9, is all that is needed.

Consider four points and the way in which the splits of T_d can separate them. If σ separates A and B from C and D , then, by compatibility, no split of T_d can separate A and C from B and D nor can it separate A and D from B and C . We can write: λ_{AB} for the sum of the μ_σ for which σ separates A and B from C and D ; λ_A for the sum of the μ_σ for which σ separates A from B, C and D ; and λ_B, λ_C and λ_D similarly. Then $\Delta_d(A, B) + \Delta_d(C, D) = \lambda_A + \lambda_B + \lambda_C + \lambda_D$ and $\Delta_d(A, C) + \Delta_d(B, D) = \Delta_d(B, C) + \Delta_d(A, D) = \lambda_A + \lambda_B + \lambda_C + \lambda_D + \lambda_{AB}$ and the four-point condition is satisfied.

For the converse, take any two members A and B of S and define a function f on S by:

$$f(X) = d(A, X) - d(B, X) \quad .$$

d is necessarily a pseudo-metric and so for all X in S $-d(A, B) \leq f(X) \leq d(A, B)$. Suppose that α and α' are real numbers within this range such that $\alpha' > \alpha$ and for no X does $f(X)$ lie between α and α' . Then the pair of subsets,

$$\{X: f(X) \geq \alpha'\} \quad \text{and} \quad \{X: f(X) \leq \alpha\}$$

forms a split. We shall show that for this split $\mu_\sigma \geq \frac{1}{2}(\alpha' - \alpha)$. By the definition of μ_σ there are members X, Y, Z, T of S such that $\mu_\sigma = \frac{1}{2}(d(X, Z) + d(Y, T) - d(X, Y) - d(Z, T))$ and such that X and Y are separated by this split from Z and T . We can assume for convenience that $f(X) \leq f(Y) \leq \alpha < \alpha' \leq f(Z) \leq f(T)$. Thus, for example, $d(X, A) + d(Z, B) < d(X, B) + d(Z, A)$ and by applying the four-point condition:

$$d(X, Z) = d(X, B) + d(Z, A) - d(A, B) \quad .$$

By further applications of the four-point condition we find:

$$d(Y, T) = d(Y, B) + d(T, A) - d(A, B) \quad ,$$

$$d(X, Y) \leq d(X, B) + d(Y, A) - d(A, B) \quad ,$$

$$d(Z, T) \leq d(Z, B) + d(T, A) - d(A, B) \quad .$$

Combining these last four expressions gives:

$$\mu_\sigma \geq \frac{1}{2}(f(Z) - f(Y)) \geq \frac{1}{2}(\alpha' - \alpha) \quad .$$

Now let $\alpha_1, \alpha_2, \dots, \alpha_p$ be the values of $f(X)$ arranged in ascending order so

that $\alpha_1 = -d(A, B) = f(A)$ and $\alpha_p = d(A, B) = f(B)$. Each pair α_i, α_{i+1} gives us a split σ_i in the manner just indicated and $\mu_{\sigma_i} \geq \frac{1}{2}(\alpha_{i+1} - \alpha_i)$. Since each σ_i separates A from B , we get:

$$\Delta_d(A, B) \geq \sum_{i=1}^{p-1} \mu_{\sigma_i} \geq \frac{1}{2} \sum_{i=1}^{p-1} (\alpha_{i+1} - \alpha_i) = d(A, B)$$

which is what we wanted to prove.

To summarize: we have defined the transformation $d \rightarrow \Delta_d$; we have shown that it defines a unique tree and preserves an additive tree metric; we have also established a necessary and sufficient condition for a DC to be an additive tree metric. One further observation will be of use to us in the discussion that follows. The transformation $d \rightarrow \Delta_d$ is continuous with respect to the natural topology on the set of DCs. Informally this means that, given reasonable definition of stress between DCs, any sufficiently small perturbation of d will result in a small perturbation of Δ_d .

DISCUSSION

Δ_N was described in the introduction as the nearest additive tree metric to d for some given definition of stress. We now turn to a practical comparison of the two transformations $d \rightarrow \Delta_d$ and $d \rightarrow \Delta_N$. Any discussion of why a particular set of data should give rise to an additive tree metric has been deliberately avoided. The justification would usually be given by some theory such as a postulate about evolutionary rates. Such a theory might also be adequate to determine how to recover the additive tree metric, but it is felt that none of the existing theories which presently figure in any of the situations mentioned is sufficient to do this. We therefore list some criticisms of Δ_N and Δ_d which may be of value in deciding which is best suited to a particular situation.

(1) Δ_N is not well defined. There may be two additive tree metrics equally near to d .

(2) There is no practical algorithm which reliably finds Δ_N on a large base set.

(3) The transformation $d \rightarrow \Delta_N$ is not continuous.

(4) Δ_d can give uninteresting trees from a DC which does not resemble an additive tree metric.

The first of these is not a pedantic quibble. DCs constructed, for example, from amino acid sequences can take on small integral values, and it is quite possible to find two additive tree metrics equally close to this DC. As for the second point: we have seen some of the difficulties involved in finding Δ_N ; these do not apply to Δ_d . It happens that there is a reasonably fast algorithm for finding Δ_d which does not involve a search through all the possible splits of S , and it is hoped to publish details of this algorithm and its implementation shortly. Whether or not the third criticism is important depends on one's prior assumptions of what form the result can take. If there are no such assumptions then it certainly is important; but there may be circumstances where, because of some appreciation of the data, one can reject alternative trees that can result from small perturbations of the original DC. Nevertheless, it is felt that continuity is a desirable property of any transformation from a

DC to an additive tree metric. The fourth criticism is very much bound up with the third. The possibility of getting uninteresting trees (that is, trees with few splits) is the price paid for continuity. A detailed tree produced by Δ_N would usually be of value either because the stress with the original DC is very low or because, having produced it, one can find some other justification for its form. In the absence of any such justification, an intricate tree may be more misleading than helpful. Moreover if Δ_N does give very low stress with d , Δ_d will do so also for much less computation.

The first and third of these criticisms are the same as those made by Jardine, Jardine, and Sibson (1967) with respect to certain methods in cluster analysis. We have also noticed a correspondence between the four-point condition and the ultrametric inequality, which is central to nearly all types of cluster analysis. These similarities are not all fortuitous. There is an extension of the theory contained in this paper which embraces both. Loosely speaking, this extension involves an asymmetry in the definition of a split and asymmetry in the derivation of a DC from a split. In the limit of lop-sidedness, one component of each split gets entirely neglected and the other turns into one of Jardine's ball clusters (Jardine 1969). The four-point condition, or rather its extension, turns into the ultrametric inequality. The case we have dealt with is the symmetrical case in this extended theory and it may be that the intermediate cases will prove of some value as well.

Computing a DC is not the only way open to us for finding a tree. For protein chains one can avoid a DC and define a set of splits in terms of the amino-acid sequences themselves; and these splits turn out to be compatible. The same thing can be done with manuscripts or for any data in which one has recorded a set of discrete attribute values for the given objects. Trees constructed in this way can be more representative of the raw data since, as we have noted, a DC can obscure useful information. One way of seeing this is that a DC is calculated on the properties of *pairs* of objects, while compatible splits can be defined by properties of objects taken *four* at a time. It is not surprising that by avoiding a DC, one can build trees which give much better descriptions of the data. Finding trees, and possibly clusters, from raw attribute data is something that deserves further investigation.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Professor H.C. Longuet-Higgins, FRS, for his good advice and constant encouragement. He would also like to thank Professor A. McIntosh, Mr C. Stephenson, Dr N. Jardine, and Dr R. Sibson for many useful discussions. This work was supported by a grant from the Royal Society.

REFERENCES

- Cavalli-Sforza, L.L. & Edwards, A.W.F. (1967) *Am. J. hum. Genet.*, **19**, 233-57.
Eck, R.V. & Dayhoff, M.O. (1966) *Atlas of Protein Sequence and Structure* 1966, pp. 164-9. Nat. biomed. res. Found.
Harary, F. (1969) *Graph Theory*. New York: Addison-Wesley.
Jardine, N. (1969) *Nature*, **244**, 185.
Jardine, C.J., Jardine, N. & Sibson, R. (1967) *Mathematical Biosciences*, **1**, 173-9.

- Jardine, N. & Sibson, R. (1971) *Mathematical Taxonomy*. New York: Wiley.
- Maas, P. (1958) *Textual Criticism* (tr. Flower, B.). Oxford University Press.
- Sokal, R. R. & Sneath, P. H. A. (1963) *Principles of Numerical Taxonomy*. W. H. Freeman.