Représentations en arbre de proximités relatives

Tree representations of qualitative proximities

Luigi Burigana

Electronic version

URL: <http://journals.openedition.org/msh/11000>

DOI: 10.4000/msh.11000

ISSN: 1950-6821

Publisher

Centre d'analyse et de mathématique sociales de l'EHESS

Printed version

Date of publication: 1 May 2009

Number of pages: 5-36

ISSN: 0987-6936

Electronic reference

Luigi Burigana, « Représentations en arbre de proximités relatives », *Mathématiques et sciences humaines* [Online], 185 | Printemps 2009, Online since 15 May 2009, connection on 23 July 2020. URL : <http://journals.openedition.org/msh/11000> ; DOI : <https://doi.org/10.4000/msh.11000>

TREE REPRESENTATIONS OF BETWEENNESS RELATIONS DEFINED BY INTERSECTION AND INCLUSION

Luigi BURIGANA¹

RÉSUMÉ – Représentations en arbre de relations d’interposition définie par intersection et inclusion

Sur une famille d’ensembles, une relation ternaire peut être définie en convenant que, pour U, V, W membres de la famille, V est considéré comme “interposé” entre U et W au cas où V inclurait l’intersection entre U et W . Cette relation est dénommée “interposition selon intersection” et elle peut être interprétée comme la description des rapports de proximité entre objets associés aux ensembles dans la famille. L’éventuel usage d’un graphe en arbre pour la représentation d’une telle relation est examiné. Des caractérisations sont démontrées aussi bien pour une représentation pleine (il existe un arbre dont l’interposition coïncide avec l’interposition selon intersection donnée : Section 2), que pour une représentation partielle (il existe un arbre dont l’interposition est incluse dans l’interposition selon intersection donnée : Section 3). Dans la Section 4 sont illustrées des procédures servant à la construction effective de solutions du problème de représentation, pleine et partielle. Dans la Section 5 sont rappelés certains paradigmes de la psychométrie moderne afin de mettre en évidence les particularités de la méthode proposée.

MOTS CLÉS – Arbre, Interposition, Proximité, Représentation

SUMMARY – *On a family of sets, a ternary relation may be defined by stating that, for U, V, W members of the family, V is “between” U and W if and only if V includes the intersection of U and W . The relation is called “intersection-betweenness” and may be understood as the description of proximities between objects associated with sets in the family. The problem of using a tree graph for representing such a relation is discussed. Characterisations are proven both for full tree representation (there is a tree-betweenness identical to the given intersection-betweenness: Section 2) and for partial tree representation (there is a tree-betweenness included in the given intersection-betweenness: Section 3). Procedures for actually finding solutions to full and partial tree representation problems are illustrated in Section 4. In Section 5 some related paradigms of modern psychometrics are mentioned, to highlight the peculiar aspects of the proposed approach.*

KEYWORDS – Betweenness, Proximity, Representation, Tree

1. INTRODUCTION

The empirical data forming the basis of analysis in this paper is some relation R between two sets X and Y . It could be, for example, that X is a set of objects, Y a set of attributes, and xRy (for any $x \in X$ and $y \in Y$) means that object x is

¹Università di Padova, Dipartimento di Psicologia Generale, Via Venezia 8, I-35131 Padova, Italy, e-mail : luigi.burigana@unipd.it

endowed with attribute y ; or that X is a set of dichotomous items in a questionnaire, Y a sample of participants, and xRy means that participant y gives positive answer to item x ; or that $X = Y$ is a set of persons and xRy (for any $x \neq y \in X$) means that person x accepts person y as a possible mate for a certain activity. In its standard set-theoretic expression, relation R is a subset of Cartesian product $X \times Y$, but there are other elementary ways of describing the same data. In particular, for the purposes of our analysis, we present the data in the following form:

$$\mathcal{X} = \{xR : x \in X\}$$

where:

$$xR = \{y \in Y : xRy\}, \text{ for each } x \in X.$$

In other words, $\mathcal{X}$ is the family of images of single elements of X under relation R , each image being a subset of Y (so that $\mathcal{X}$ is a family of subsets of Y indexed by the elements of X). For simplicity of notation, if a certain lower-case letter denotes an element of X , then the corresponding upper-case letter denotes the image of that element under relation R — thus, if u, v, w denote elements of X , then U, V, W denote images uR, vR, wR , which are subsets of Y .

The theoretical analysis presented in this paper is not novel in referring to a relation between two sets as the basic empirical data, and in referring to such a relation for representational purposes (approaches sharing these premises are mentioned in Section 5 for comparison). However, there are two elementary features distinguishing the perspective of this study, one concerning the information to be extracted for representation, and the other fixing the general way in which that representation must be accomplished.

The information to be extracted is a ternary relation, expressing “qualitative proximity” on a set-theoretic basis.

DEFINITION 1. *Let X and Y be sets, and $\mathcal{X}$ a family of subsets of Y indexed by elements of X (for each $u \in X$, $U \in \mathcal{X}$ is the subset of Y indexed by u). The intersection-betweenness induced by $\mathcal{X}$ on X , denoted by $B_{\mathcal{X}}$, is the set of triples (u, v, w) of distinct elements of X such that $U \cap W \subseteq V$.*

For example, if X and Y are a set of objects and a set of attributes, then $(u, v, w) \in B_{\mathcal{X}}$ means that $U \cap W$ (the set of attributes shared by objects u and w) is a subset of both $U \cap V$ (the set of attributes shared by u and v) and $V \cap W$ (the set of attributes shared by v and w). Thus, if we take the set of attributes shared by any two objects as an index of the “qualitative proximity” of the objects themselves, then $(u, v, w) \in B_{\mathcal{X}}$ means that the proximities of v to u and to w are closer than the proximity of u to w , which may be described by saying that v lies “between” u and w . For brevity of notation, we write uvw to mean $(u, v, w) \in B_{\mathcal{X}}$, so that, in the presumed context:

$$\begin{aligned} uvw &\text{ iff } U \cap W \subseteq V \\ &\text{ iff } U \cap V^c \cap W = \emptyset. \end{aligned} \tag{1}$$

This set-theoretic way of determining a ternary (betweenness) relation is not a novelty in psychometrics. In particular, this definition, combined with the assumption

of a linear order, was used by Tversky and Gati to specify the concept of a “chain”, as understood within the frame of their “contrast model” of similarity relations (cf. [Tversky, Gati, 1982, p. 129; Gati, Tversky, 1982, p. 329; Candel, 1997]). In the concluding section I point out what is new in our results concerning intersection-betweenness, compared with the position of the same concept in Tversky and Gati’s theorisation.

As models for representation we refer to tree graphs.

DEFINITION 2. *Let A be a connected and acyclic adjacency on set X – so that (X, A) is a tree. The tree-betweenness induced by A on X , denoted by B_A , is the set of all triples (u, v, w) of distinct elements of X such that v is a vertex in the path joining u and w within the tree.*

The use of tree graphs for representation purposes – in particular, for representation of proximity or similarity data – is a conspicuous part of modern psychometrics (cf. [Barthélemy, Guénoche, 1991], for systematic treatment of the matter). Again, I postpone to the last section some comments on the peculiarities of our reference to tree graphs, compared with well-established uses of such graphs in psychometrics.

We are interested in two possible relationships between intersection-betweenness and tree-betweenness on the same domain (one relationship is tighter than the other).

DEFINITION 3. *Let us presume a context characterised by all assumptions in the definitions above. Tree (X, A) makes a full representation or a partial representation of intersection-betweenness B_X , depending on whether tree-betweenness B_A is identical to or included in B_X .*

If we denote by $X^{(3)}$ the set of all triples of distinct elements of X and, besides simplifying “ $(u, v, w) \in B_X$ ” into “ uvw ”, also simplify “ $(u, v, w) \in B_A$ ” into “ $u.v.w$ ”, then the conditions defining full and partial tree representations are as follows:

$$\begin{aligned} &u.v.w \text{ if and only if } uvw, \text{ for all } (u, v, w) \in X^{(3)} \\ &\text{if } u.v.w, \text{ then } uvw, \text{ for all } (u, v, w) \in X^{(3)}. \end{aligned}$$

The main object of this paper is to present characterisations both of full and of partial tree representations of intersection-betweennesses. This is done in Sections 2 and 3, by applying concepts from the algebraic theory of betweenness relations (for the full representation problem) and from the combinatorial theory of relational databases (for the partial representation problem). In defining and proving the characterisations, we consider chains of operations which may be viewed as procedures for actually constructing solutions to the tree representation problems – when the problems are solvable, of course. These constructive procedures are illustrated in Section 4 with two examples. Lastly, in Section 5 reference is made to some notable topics of combinatorial psychometrics which have intersections with the subject of this study, either because they refer to a relation between two sets as original data, or because they make use of trees for representation purposes.

2. CONDITIONS FOR FULL TREE REPRESENTATION

There are properties which are “intrinsic” to intersection-betweenness, in that they are directly implied by the formula defining the concept (equation (1)) and by regularities of the set-theoretic operations involved in the formula. Intrinsic properties are true of any intersection-betweenness, regardless of peculiarities of the family of sets inducing it. Three such properties are those given by the following formulas (in these and similar equations, variables $u, v, w, \dots$ are presumed to be universally quantified, i.e., opening “for all $u, v, w, \dots$ in domain X ” is understood, unless otherwise stated):

$$\text{if } uvw, \text{ then } wvu \tag{2}$$

$$\text{if } uvw \text{ and } uzv, \text{ then } uzv \tag{3}$$

$$\text{if } tuz \text{ and } twz \text{ and } uvw, \text{ then } tvz.$$

The first is a symmetry property, and is directly implied by the commutative law of intersection. The second is a form of transitivity on four elements, and is proven true by considering that, if $U \cap V^c \cap W = \emptyset$ and $U \cap V \cap Z^c = \emptyset$, then $U \cap W \cap Z^c = (U \cap V \cap W \cap Z^c) \cup (U \cap V^c \cap W \cap Z^c) = \emptyset \cup \emptyset = \emptyset$. The third is a kind of transitivity on five elements and is derived by means of a similar argument. (The identification and comparison of several possible “transitivities” of a ternary relation, involving four, five, or more elements in the domain, is a typical subject of the algebraic theory of betweenness relations; cf. [Pitcher, Smiley, 1942]).

There are several other properties which are *not* intrinsic to intersection-betweenness, in the sense above: each of them may be true or false of any given intersection-betweenness, depending on the peculiarities of the family of sets inducing that betweenness. Of critical importance for the analysis to be done are the following four properties:

$$\text{if } uvw, \text{ then not}(vuw) \tag{4}$$

$$\text{if } uvw \text{ and } vwz, \text{ then } uwz \tag{5}$$

$$\text{if } uvw \text{ and } uwz, \text{ then } vwz \tag{6}$$

$$\text{if } N(u, v, w), \text{ then there exists } c \text{ so that } ucv \text{ and } ucw \text{ and } vcw \tag{7}$$

where writing $N(u, v, w)$ means $\text{not}(uvw \text{ or } vwu \text{ or } wuv)$ and $u \neq v \neq w \neq u$. Of these formulas, the first expresses an asymmetry of a ternary relation, the second and the third specify two forms of transitivity of such a relation, and the fourth describes a kind of conditional completeness (diagrams illustrating (5), (6) and (7) are shown in Figure 1). By referring to any intersection-betweenness, as induced by some family of sets $\mathcal{X}$ through (1), it is easily seen that property (4) holds true if and only if the following set-theoretic condition is satisfied:

$$(U + V) \cap W \neq \emptyset, \tag{8}$$

where $+$ means symmetric difference (i.e., $U + V = U \setminus V \cup V \setminus U = (U \cap V^c) \cup (U^c \cap V)$); properties (5) and (6) are jointly true (i.e., the statement “for all u, v, w, z , if uvw ,

then $(uwz \text{ iff } vwz)$ ” is true) if and only if the following set-theoretic condition is satisfied:

$$\text{if } U \cap V^c \cap W = \emptyset, \text{ then } (U \cap W^c \cap Z = \emptyset \text{ iff } V \cap W^c \cap Z = \emptyset); \quad (9)$$

and property (7) holds true if and only if the next set-theoretic condition is satisfied:

$$\begin{aligned} &\text{if } U^c \cap V \cap W \neq \emptyset \text{ and } U \cap V^c \cap W \neq \emptyset \text{ and } U \cap V \cap W^c \neq \emptyset, \\ &\text{then there is } C \in \mathcal{X} \text{ so that } C \supseteq U \cap V \cap W. \end{aligned} \quad (10)$$

Moreover, it is shown that each of the four properties (4)–(7) is independent of the other three, i.e., there are situations in which that property is false, whereas the other three are true. The following families of sets are simple examples to this effect. Specifically, the intersection-betweenness induced by the first family falsifies (4) but verifies (5)–(7) (in the empty mode), and thus proves the independence of (4); similarly, the second, third and fourth families prove the independence of (5), (6) and (7), respectively.

$$\begin{aligned} &\{\{a\}, \{b\}, \{c\}, \{d\}\} \\ &\{\{a, b\}, \{b, c\}, \{c, d\}, \{a, d\}\} \\ &\{\{a, b, c, d\}, \{a, b\}, \{b, c\}, \{c, d\}\} \\ &\{\{a, b, d, e\}, \{a, b, c\}, \{b, c, d\}, \{c, d, e\}\}. \end{aligned}$$

Note that transitivities (5) and (6) (which are not universally true of intersection-betweennesses) jointly imply transitivity (3) (which is instead universally true of intersection-betweennesses). Actually, uvw and uzv imply zvw (by presuming (6)), and zvw and uzv imply uzw (by presuming (5)).

Conditions (2) and (4)–(7) make a notable system: they jointly characterise tree-betweenness, as we shall show by proving Theorem 1. This result is not completely new: in the literature other characterisations of tree-betweenness may be found, e.g., Sholander [1952] and Defays [1979]². Besides some differences in the axioms, what primarily distinguishes our proposal is the constructive character of the proof:

²In Sholander [1952], the characterisation of tree-betweenness is a part of a general analysis, concerning not only trees, but also orders and lattices. The main constructs are a system $S \subseteq 2^X$ of subsets of a basic set X (which are indexed by pairs of elements of X , and are called “segments”) and a ternary relation $B \subseteq X^3$ (called “betweenness”). Moreover, a simple rule linking both constructs is defined: for all $u, v, w \in X$, there is $[u, w] \in S$ so that $v \in [u, w]$, if and only if $(u, v, w) \in B$. By this rule, any system of segments S determines a betweenness relation $b(S)$ and, vice-versa, any betweenness relation B determines a system of segments $s(B)$, so that $s(b(S)) = S$ and $b(s(B)) = B$. With reference to a system of segments S , the following conditions are considered (which jointly define the concept of a “tree”, as understood by Sholander): (S) for all $[u, v], [v, w] \in S$ there is $[v, x] \in S$ so that $[u, v] \cap [v, w] = [v, x]$; (T) for all $[u, v], [u, w] \in S$, if $[u, v] \subseteq [u, w]$, then $[u, v] \cap [v, w] = \{v\}$; (U₁) for all $[u, v], [v, w] \in S$, if $[u, v] \cap [v, w] = \{v\}$, then $[u, v] \cup [v, w] = [u, w]$. With reference to a betweenness relation B , the following conditions are considered: (B) for all $u, v \in X$, $(u, v, u) \in B$ iff $u = v$; (C) for all $u, v, w, x, y \in X$, if $(u, v, w), (v, x, y) \in B$, then $(w, v, x) \in B$ or $(y, v, u) \in B$; (D₁) for all $u, v, w \in X$, there is $x \in X$ so that $(u, x, v), (v, x, w), (w, x, u) \in B$. What Sholander shows is that a system of segments S satisfies conditions (S), (T) and (U₁) – i.e., S is a “tree” – if and only if the associated betweenness $b(S)$ satisfies conditions (B), (C) and (D₁). The characterisation by Defays [1979] basically coincides with our own in its first four axioms, but differs in the fifth, which is the following: for all $u, v \in X$,

FIGURE 1. Diagrams of implications (5), (6) and (7). Link uvw – i.e., $(u, v, w) \in B$ – is a two-tail arrow pointing to v . Condition $N(u, v, w)$ – i.e., $(u, v, w) \notin B$ and $(v, w, u) \notin B$ and $(w, u, v) \notin B$ – is a cycle of three dotted arcs. Any sequence $(u_1, \dots, u_m)$ so that $u_i u_j u_k$ for all $1 \leq i < j < k \leq m$ is represented as a chain. Any quadruple $\{u, v, w, c\}$ so that ucv and ucw and vcw is a three-ray star centred on c .

we prove part “if” of Theorem 1 by defining an inductive procedure through which, based on a ternary relation B complying with the stated conditions, a monotone sequence of trees can be constructed so that the last tree in the sequence has precisely B as its betweenness relation. That procedure will be illustrated in Section 4 (first example).

We prepare the proof of the theorem by deriving some properties from conditions (2) and (4)–(7). They are expressed as formulas in separate lemmas, and illustrated by diagrams in the way we used in Figure 1.

LEMMA 1. *Let B be any ternary relation complying with conditions (2), (4)–(7) on a domain X , and N its complementary relation – i.e., $N(u, v, w)$ means not(uvw or vwu or wuv) and $u \neq v \neq w \neq u$. Then, for all $t, u, v, w, x, z \in X$ such that t, u, v, w are distinct from one another and from x and z :*

if $(u, x, v) \notin B$ for all $x \in X \setminus \{u, v\}$, then $((x, u, v) \in B$ or $(u, v, x) \in B)$ for all $x \in X \setminus \{u, v\}$. Defays proves part “if” of his main theorem by first deriving a binary relation A from ternary relation B so that, for all $u, v \in X$: uAv iff $(u, x, v) \notin B$ for all $x \in X \setminus \{u, v\}$; and then showing that, if B satisfies the five stated conditions, then A is acyclic and connected, and the betweenness it induces on X is precisely B .

- (i) if tuw and tvw , then either $(tuv$ and $uvw)$ or $(tvu$ and $vuw)$;
- (ii) if uxv , uxw , and vwx , then $N(u, v, w)$;
- (iii) if uxv , uxw , vwx and uzv , uzw , vzw , then $x = z$.

Proof. – Part (i): Hypotheses tuw and tvw imply $\text{not}(uvw)$ (due to (2), (4) and (5)). The hypotheses also imply $\text{not}(N(u, v, w))$, because otherwise $c \in X \setminus \{u, v, w\}$ would exist so that ucv and ucw and vcw (due to (7)), but (tuw, ucv, ucw) implies tuv (due to (5) and (6)), and similarly (tvw, vcu, vcw) implies tvu , thus contradicting (4). Thus, in view of (2) and (4), we must have either uvw or vuw and, for similar reasons, either tuv or tvu . But tuv combined with hypothesis tvw implies uvw (due to (6)), which is incompatible with vuw (due to (4)), and for the same reasons tvu is incompatible with uvw . Thus, either $(tuv$ and $uvw)$ or $(tvu$ and $vuw)$, which is the consequent in the stated implication.

– Part (ii): Suppose uxv , uxw , and vwx . Then $\text{not}(uvw)$, because otherwise xvw (due to hypothesis uxv and (6)), which contradicts hypothesis vwx (in view of (4)). For similar reasons, $\text{not}(vwu)$ and $\text{not}(wuv)$, so that $N(u, v, w)$.

– Part (iii): Presume (uxv, uxw, vwx) and (uzv, uzv, vzw) , and (to reach a contradiction) $x \neq z$. Hypotheses uxv , uzv , and $x \neq z$ imply that either uxz or uzx , due to part (i) proven above. By applying (2) and (6) at each step, we see that alternative uxz implies vzx (due to uzv), which implies wxz (due to vwx), which implies uzx (due to uzw), which contradicts uxz (in view of (4)). The same contradiction is obtained starting from alternative uzx . $\square$

Part (iii) of the proven lemma reinforces property (7): it ensures that, in the presumed conditions, if $\{u, v, w\}$ is a triple of elements so that $N(u, v, w)$, then there is *only one* element c so that ucv , ucw , and vcw . We denote this element by $c(u, v, w)$, and call it the *median* of triple $\{u, v, w\}$ (this is the standard name for similarly defined concepts; cf. [Sholander, 1952, §5; Barthélemy, Guénoche, 1991, §2.4.2].

LEMMA 2. *Let B be any ternary relation satisfying conditions (2), (4)–(7) on a domain X , and N its complementary relation. Let t, u, v, w be distinct elements of X so that utw and $N(u, v, w)$. Then exactly one of these three circumstances must be true:*

- (i) utv , $N(t, v, w)$, $c(t, v, w) = c(u, v, w)$;
- (ii) vtw , $N(t, u, v)$, $c(t, u, v) = c(u, v, w)$;
- (iii) $c(u, v, w) = t$.

Proof. The mutual incompatibility of the three circumstances is obvious (consider, in particular, that (iii) implies utv and utw and vtw). The exhaustiveness of the

three circumstances — i.e., that at least one of them must be true — can be proven by showing that $N(t, v, w)$ implies the other two statements in (i), $N(t, u, v)$ implies the other two statements in (ii), and $\text{not}(N(t, v, w) \text{ or } N(t, u, v))$ implies (iii).

– First task: Presume $N(t, v, w)$ and let x be the median $c(t, v, w)$, so that txv , txw , and vxw . Hypothesis utw combined with txw implies uxw and utx (due to (3) and (6)); in turn, utx and txv imply uxv (due to (5)); thus, uxv and uxw and vxw , so that $c(u, v, w) = x = c(t, v, w)$ (because of Lemma 1). Moreover, txv and utx imply utv (due to (5)).

– Second task: The implication concerning situation (ii) is proven in the same manner.

– Third task: Presume $\text{not}(N(t, v, w))$, which means tvw or vwt or wtv . The first disjunct combined with hypothesis utw implies uvw (due to (3)), contradicting hypothesis $N(u, v, w)$, so that it must be discarded. The impossibility of the second disjunct is proven through a similar argument. Thus, by exclusion, it must be wtv . In the same manner, it is proven that $\text{not}(N(t, u, v))$ implies utv . Hence, utw (by hypothesis) and wtv and utv , which means $t = c(u, v, w)$ (according to Lemma 1). $\square$

Given a ternary relation B , if $(u_1, \dots, u_m)$ is a sequence of $m \geq 3$ elements in its domain so that $u_i u_j u_k$ (i.e., $(u_i, u_j, u_k) \in B$) for all $1 \leq i < j < k \leq m$, then we call it a B -chain, and write $u_1 u_2 \dots u_m$. For example, using this notation, both disjuncts forming the consequent of implication (i) in Lemma 1 would be written as $tuvw$ and $tvuw$.

LEMMA 3. Let B be any ternary relation satisfying conditions (2), (4)–(7) on a domain X , N its complementary relation, $(u_1, \dots, u_m)$ and $(v_1, \dots, v_n)$ two B -chains in X , and w any element in $X \setminus \{u_1, \dots, u_m\}$.

- (i) If $u_1 w u_m$, then $u_i w u_{i+1}$ for some $i \in \{1, \dots, m-1\}$.
- (ii) If $N(u_1, w, u_m)$, then either $N(u_i, w, u_{i+1})$ for some $i \in \{1, \dots, m-1\}$, or $c(u_i, w, u_{i+2}) = u_{i+1}$ for some $i \in \{1, \dots, m-2\}$.
- (iii) If $u_{m-1} = v_1$, $u_m = v_2$, and $\{u_1, \dots, u_m\} \cap \{v_3, \dots, v_n\} = \emptyset$, then $(u_1, \dots, u_m, v_3, \dots, v_n)$ is a B -chain.
- (iv) If there is $h \in \{1, \dots, m-1\}$ so that $u_h = v_1$, $u_{h+1} = v_n$ and $\{u_1, \dots, u_m\} \cap \{v_2, \dots, v_{n-1}\} = \emptyset$, then $(u_1, \dots, u_h, v_2, \dots, v_{n-1}, u_{h+1}, \dots, u_m)$ is a B -chain.

Proof.

– Part (i) is proven by induction on length m of the B -chain and by applying Lemma 1.i.

– Part (ii): First we prove that, in the presumed conditions, the following implication is true:

$$\text{if } N(u_1, w, u_m), \text{ then } N(u_i, w, u_{i+2}) \text{ for some } i \in \{1, \dots, m-2\} \quad (11)$$

which means:

$$\text{if not}(N(u_i, w, u_{i+2})) \text{ for all } i \in \{1, \dots, m-2\}, \text{ then not}(N(u_1, w, u_m)).$$

The implication is obviously true for $m = 3$; let us thus presume $m \geq 4$. If wu_1u_3 , then wu_1u_m (due to hypothesis $u_1u_3u_m$ and (5)), so that $\text{not}(N(u_1, w, u_m))$. The same conclusion follows from $u_{m-2}u_mw$. Moreover, if u_iwu_{i+2} for some $i \in \{1, \dots, m-2\}$, then u_1wu_m (due to hypotheses $u_1u_iu_{i+2}$ and $u_1u_{i+2}u_m$, and (3)), so that, again, $\text{not}(N(u_1, w, u_m))$. Given these results, to complete the proof of (11), it suffices to show that the conjunction of the following four statements is false, in the presumed conditions:

$$\begin{aligned} &\text{not}(N(u_i, w, u_{i+2})) \text{ for all } i \in \{1, \dots, m-2\}; \quad \text{not}(wu_1u_3); \\ &\text{not}(u_{m-2}u_mu_mw); \quad \text{not}(u_iwu_{i+2}) \text{ for all } i \in \{1, \dots, m-2\}. \end{aligned}$$

The first and fourth statements imply that, for all $i \in \{1, \dots, m-2\}$, either wu_iu_{i+2} or $u_iu_{i+2}w$. Based on this, by defining $k = \max\{i \in \{1, \dots, m-2\} : u_iu_{i+2}w\}$, we obtain that the second and third statements imply $1 \leq k \leq m-3$, so that $u_ku_{k+2}w$ and $wu_{k+1}u_{k+3}$. But these two results, when combined with hypotheses $u_ku_{k+1}u_{k+2}$ and $u_{k+1}u_{k+2}u_{k+3}$, imply $u_{k+1}u_{k+2}w$ and $wu_{k+1}u_{k+2}$ (due to (6)), which is impossible in view of (4), and this completes the proof of (11). Now, if $N(u_1, w, u_m)$, then $N(u_i, w, u_{i+2})$ for some $i \in \{1, \dots, m-2\}$ (due to (11)) and $u_iu_{i+1}u_{i+2}$ (because the sequence is a B -chain), so that, by applying Lemma 2 (after substituting w, u_i, u_{i+1}, u_{i+2} for v, u, t, w), we obtain the two disjuncts in part (ii) of this lemma (the first disjunct corresponds to cases (i) and (ii) in Lemma 2, the second disjunct to case (iii)).

– Part (iii) is proven by first showing that, in the presumed conditions, sequence $(u_1, \dots, u_{m-1}, u_m, v_3)$ is a B -chain, i.e., $u_iu_jv_3$ for all $1 \leq i < j \leq m$. For $i = m-1$ and $j = m$, statement $u_iu_jv_3$ is obviously true, because of hypotheses $u_{m-1} = v_1$,

$u_m = v_2$, and $v_1v_2v_3$. For $i < m - 1$ and $j = m$, statement $u_iu_jv_3$ derives from hypotheses $u_iu_{m-1}u_m$ and $v_1v_2v_3$ (due to (5)). For $i < j < m$, statement $u_iu_jv_3$ is implied by $u_ju_mv_3$ and hypothesis $u_iu_ju_m$ (again due to (5)). By inductively applying this argument, it is seen that sequences $(u_1, \dots, u_m, v_3, v_4), \dots, (u_1, \dots, u_m, v_3, v_4, \dots, v_n)$ are all B -chains.

– Part (iv) is proven in a similar way, by appealing to (3) (instead of (5)). $\square$

Let B be a ternary relation on a set X , Z a subset of X , and A a binary relation on Z which is symmetric, acyclic and connected, so that (Z, A) is a tree. Structure (Z, A) is called a B -tree if $B_A \subseteq B$, where B_A is the tree-betweenness induced on Z by A (Definition 2). Note that, by using this terminology, we may say that finding a “partial tree representation” of ternary relation B (Definition 3) means finding a B -tree the domain of which is whole set X .

LEMMA 4. *Let B be any ternary relation complying with conditions (2), (4)–(7) on a domain X , N its complementary relation, (Z, A) a B -tree, and u, v, w distinct elements of Z so that $uAvAw$. Then, for each $x \in X \setminus Z$, the following implication holds true:*

$$\text{if } (vxu \text{ or } vux \text{ or } N(v, u, x)), \text{ then not}(vxw \text{ or } vwx \text{ or } N(v, w, x)).$$

Proof. First, consider that hypotheses $uAvAw$ and $B_A \subseteq B$ imply uvw . To prove the stated implication means proving that, in the presumed conditions, each of the nine conjunctions between disjuncts in the antecedent and in the consequent is impossible – actually, $(p_1 \vee p_2 \vee p_3) \Rightarrow \neg(q_1 \vee q_2 \vee q_3)$ and $\neg((p_1 \vee p_2 \vee p_3) \wedge (q_1 \vee q_2 \vee q_3))$ are equivalent formulas of propositional calculus. Because of symmetries, only six conjunctions need be discussed.

– Case vxu and vxw : Link uxw would imply $N(u, v, w)$ (due to Lemma 1.ii), contradicting uvw . Link xuw would imply vuw (due to (5)), again contradicting uvw (in view of (4)). Link uwx must be discarded for similar reasons. Lastly, if $N(u, x, w)$, then utx and xtw and utw (with $t = c(u, x, w)$), so that vtu and vtw (due to (3)), which implies $N(u, v, w)$ (by Lemma 1.ii), again contradicting uvw .

– Case vxu and vwx : Due to (3), this implies vwu , which contrasts with uvw (because of (4)).

– Case vxu and $N(v, w, x)$: This would imply vtw and vtx and wtx , with $t = c(v, w, x)$. But vxu and vtx imply vtu and txu (due to (3) and (6)), and txu combined with wtx implies wtu (due to (5)), so that utv and utw and vtw , which, by Lemma 1.ii, implies $N(u, v, w)$, thus contradicting uvw .

– Case vux and vwx : According to Lemma 1.i it should be either vuw or vwu , which are both incompatible with uvw , in view of (4).

– Case vux and $N(v, w, x)$: This would imply vtw and vtx and xtw , with $t = c(v, w, x)$. Because of Lemma 1.i, vux and vtx imply either vut or $(vtu$ and $tux)$. In the former case, we would obtain vuw (because vtw and by (3)), in contrast with uvw . In the latter case, we would obtain vtw and vtu and utw (which is implied by xtw and tux , due to (6)), which implies $N(u, v, w)$ (by Lemma 1.ii), again in contrast with uvw .

– Case $N(v, u, x)$ and $N(v, w, x)$: If both statements were true, then vtu , vtx , and utx with $t = c(v, u, x)$, and vzw , vzx , and wzx with $z = c(v, w, x)$. Suppose $t = z$.

Then hypothesis uvw is contradicted by uwt (when combined with vzw , due to (5) and (4)), by wtu (when combined with vtu and vzw , due to Lemma 1.ii), and by tuw (when combined with vzw , due to (3) and (4)), so that $N(u, w, t)$, which implies usw , ust , and wst , with $s = c(u, w, t)$. But vtu and ust imply vsu (due to (3)), vzw (i.e., vtw) and wst imply vsu for the same reason, and usw is presumed, so that $N(u, v, w)$ by Lemma 1.ii, which contradicts hypothesis uvw . Then suppose $t \neq z$. Links vtx and vzx imply that either $vtzx$ or $vztx$, according to Lemma 1.i. B -chain $vtzx$ leads to link vtw (deriving from vtz and vzw through (3)), link utw (deriving from vtz , vzw and tzx , utx through (6) and (5)), while link vtu is presumed, so that $N(u, v, w)$ by Lemma 1.ii, thus contradicting hypothesis uvw . B -chain $vztx$ leads to the same contradiction. $\square$

We are now in a position to prove the main result of this section (a characterisation of tree-betweenness).

THEOREM 1. *Let X be a set and B a ternary relation on it ($B \subseteq X^{(3)}$). There is a tree-adjacency A on X so that $B_A = B$ (B_A being the tree-betweenness induced by A) if and only if conditions (2) and (4)–(7) are universally satisfied by triples in B .*

Proof. That the five conditions are true of any tree-betweenness is easily proven. Thus, we focus our argument on the “if” part of the statement. More precisely, we prove this implication:

if B is a ternary relation satisfying conditions (2), (4)–(7) on a set X , (12)
then a tree-adjacency A can be constructed on X so that $B_A \subseteq B$.

Note that, in the stated conditions, inclusion $B_A \subseteq B$ implies inclusion $N_A \subseteq N$, where N_A and N are the complementary relations of B_A and B , respectively. Actually, for all $\{u, v, w\} \subseteq X$, if $N_A(u, v, w)$, then $u.x.v$ and $u.x.w$ and $v.x.w$ for some $x \in X \setminus \{u, v, w\}$ (due to (7)), so that uxv and uxw and vwx (because of hypothesis $B_A \subseteq B$), which implies $N(u, v, w)$ (by Lemma 1.ii). Also note that, in the stated conditions, if uvw , then $\text{not}(N(u, v, w))$, and $\text{not}(N_A(u, v, w))$ (due to inclusion $N_A \subseteq N$), so that $u.v.w$ or $v.w.u$ or $w.u.v$; but the second and third disjuncts imply $v.w.u$ or $w.u.v$ (due to inclusion $B_A \subseteq B$), which are incompatible with hypothesis uvw (due to (4)); hence $u.v.w$ must be true. By generalisation, we obtain that, in the stated conditions, inclusion $B_A \subseteq B$ implies inclusion $B \subseteq B_A$, hence equality $B_A = B$ by antisymmetry: this ensures that, by proving (12), we really reach our goal. We prove (12) in a constructive way, i.e., by defining a method which — in the stated conditions — allows us to construct a series of structures $T_1 = (X_1, A_1)$, $T_2 = (X_2, A_2), \dots$ so that, for each $j = 1, 2, \dots$, the following conditions are satisfied:

- α : X_j is a proper subset of X_{j+1} (and a subset of X)
- β : A_j is a tree-adjacency on X_j
- γ : betweenness B_j of tree $T_j = (X_j, A_j)$ is included in B .

Due to α and the finiteness of X , after a suitable number of steps a structure $T_n = (X_n, A_n)$ is obtained so that $X_n = X$, $A = A_n$ is a tree-adjacency (condition β),

and $B_A = B_n \subseteq B$ (condition γ): this is precisely the consequent of implication (12). The method to be described is inductive in character. The *start rule* of the method prescribes choosing any triple $\{u, v, w\} \subseteq X$ so that uvw (i.e., $(u, v, w) \in B$), and putting $X_1 = \{u, v, w\}$, $A_1 = \{\{u, v\}, \{v, w\}\}$. Hypotheses $|X| \geq 3$ and (7) ensure that a choice like this can really be made. It is obvious that structure $T_1 = (X_1, A_1)$ has properties β and γ . The *inductive rule* of the method refers to a structure $T_j = (X_j, A_j)$ with $X_j \subset X$, presumed to have been obtained at the j -th step of the process and to have properties β and γ , it prescribes choosing a sequence $(u_1, \dots, u_m)$ forming a maximal path (from leaf to leaf) in tree T_j and any point $v \in X \setminus X_j$, and it guides us in inserting the new point into the given tree using the chosen maximal path as a reference, so as to form a structure $T_{j+1} = (X_{j+1}, A_{j+1})$ which itself has properties β and γ , and is such that $X_j \subset X_{j+1}$, as $v \in X_{j+1} \setminus X_j$. As T_j has property γ , sequence $(u_1, \dots, u_m)$ is a B -chain, so that (due to (4), the definition of N , and Lemma 3) exactly one of these five cases must occur:

- I: vu_1u_2
- II: $u_{m-1}u_mv$
- III: u_ivu_{i+1} for one $i \in \{1, \dots, m-1\}$
- IV: $N(u_i, v, u_{i+1})$ for one $i \in \{1, \dots, m-1\}$
- V: $c(u_i, v, u_{i+2}) = u_{i+1}$ for one $i \in \{1, \dots, m-2\}$.

Denoting by H the neighbourhood of point u_{i+1} in tree T_j , Lemma 4 implies that, if case V is true, then exactly one of the following five sub-cases must occur:

- V.1: $|H| = 2$ (i.e., $H = \{u_i, u_{i+2}\}$)
- V.2: $|H| > 2, vu_{i+1}z$ for all $z \in H \setminus \{u_i, u_{i+2}\}$
- V.3: $|H| > 2, N(v, u_{i+1}, w)$ for one $w \in H \setminus \{u_i, u_{i+2}\}, vu_{i+1}z$ for all $z \in H \setminus \{u_i, u_{i+2}, w\}$
- V.4: $|H| > 2, u_{i+1}vw$ for one $w \in H \setminus \{u_i, u_{i+2}\}, vu_{i+1}z$ for all $z \in H \setminus \{u_i, u_{i+2}, w\}$
- V.5: $|H| > 2, u_{i+1}wv$ for one $w \in H \setminus \{u_i, u_{i+2}\}, vu_{i+1}z$ for all $z \in H \setminus \{u_i, u_{i+2}, w\}$.

Thus, the general inductive rule of the method is actually formed of nine specific rules, corresponding to cases I–IV and sub-cases V.1–V.5. The first eight rules are given in Table 1. It is easily seen that each specific rule enlarges tree-adjacency A_j

case	X_{j+1}	A_{j+1}
I	$X_j \cup \{v\}$	$A_j \cup \{\{v, u_1\}\}$
II	$X_j \cup \{v\}$	$A_j \cup \{\{u_m, v\}\}$
III	$X_j \cup \{v\}$	$(A_j \setminus \{\{u_i, u_{i+1}\}\}) \cup \{\{u_i, v\}, \{v, u_{i+1}\}\}$
IV	$X_j \cup \{v, x\}$ with $x = c(u_i, v, u_{i+1})$	$(A_j \setminus \{\{u_i, u_{i+1}\}\}) \cup \{\{u_i, x\}, \{x, u_{i+1}\}, \{x, v\}\}$
V.1	$X_j \cup \{v\}$	$A_j \cup \{\{u_{i+1}, v\}\}$
V.2	$X_j \cup \{v\}$	$A_j \cup \{\{u_{i+1}, v\}\}$
V.3	$X_j \cup \{v, x\}$ with $x = c(v, u_{i+1}, w)$	$(A_j \setminus \{\{u_{i+1}, w\}\}) \cup \{\{u_{i+1}, x\}, \{v, x\}, \{w, x\}\}$
V.4	$X_j \cup \{v\}$	$(A_j \setminus \{\{u_{i+1}, w\}\}) \cup \{\{u_{i+1}, v\}, \{v, w\}\}$

Table 1. Eight rules for enlarging tree $T_j = (X_j, A_j)$ into tree $T_{j+1} = (X_{j+1}, A_{j+1})$ through insertion of point $v \in X \setminus X_j$.

into relation A_{j+1} which itself is a tree-adjacency, i.e., property β is preserved in

passing from T_j to T_{j+1} . It is routine, though tedious, to show that the same holds true of property γ . We illustrate it by discussing the first and last rules in the table.

– Case I: Point v is the only one added in passing from tree T_j to tree T_{j+1} . It is attached as a leaf to point u_1 (which in turn is a leaf in T_j and is attached to u_2). Thus, the only new links in $B_{j+1} \setminus B_j$ are of type $v.x.y$, which means that there is a path $(w_1, w_2, w_3, \dots, w_p)$ in T_{j+1} so that $w_1 = v$, $w_2 = u_1$, $w_3 = u_2$, $w_p = y$, and $w_h = x$ for some $h \in \{2, \dots, p-1\}$. Path $(w_2, w_3, \dots, w_p)$ is inside tree T_j , hence it is a B -chain, due to property γ of T_j . As vu_1u_2 (which is the hypothesis of case I), thanks to Lemma 3.iii we infer that $(w_1, w_2, w_3, \dots, w_p)$ is itself a B -chain, which implies $w_1w_hw_p$, i.e., vxy . Through generalisation, this proves that $B_{j+1} \setminus B_j \subseteq B$, hence $B_{j+1} \subseteq B$, as $B_j \subseteq B$ due to property γ of T_j .

– Case V.4: Tree T_{j+1} is obtained from tree T_j by splitting line $\{u_{i+1}, w\}$ into lines $\{u_{i+1}, v\}$ and $\{v, w\}$, through insertion of point v . Let $T'_j = (X'_j, A'_j)$ and $T''_j = (X''_j, A''_j)$ be the two branches of T_j separated by line $\{u_{i+1}, w\}$, with $u_{i+1} \in X'_j$ and $w \in X''_j$. Any new link in $B_{j+1} \setminus B_j$ is of one of three types: $v.x.y$ with $x, y \in X'_j$, $v.x.y$ with $x, y \in X''_j$, $x.v.y$ with $x \in X'_j$ and $y \in X''_j$. The first type implies that there is a path $(w_1, w_2, \dots, w_p)$ in T_{j+1} so that $w_1 = v$, $w_2 = u_{i+1}$, $w_p = y$ and $w_h = x$ for some $h \in \{2, \dots, p-1\}$; thus, sequence $(w, w_2, \dots, w_p)$ is a path in T_j , and it is a B -chain (due to property γ of T_j); but $u_{i+1}vw$, i.e., ww_1w_2 by hypothesis, so that $(w, w_1, w_2, \dots, w_p)$ is itself a B -chain (due to Lemma 3.iv), hence $w_1w_hw_p$, i.e., vxy . A similar argument leads to the same conclusion for the second type. The third type implies that T_{j+1} contains a path $(w_1, \dots, w_{h-1}, w_h, w_{h+1}, \dots, w_p)$ so that $w_1 = x$, $w_{h-1} = u_{i+1}$, $w_h = v$, $w_{h+1} = w$, and $w_p = y$; hence $(w_1, \dots, w_{h-1}, w_{h+1}, \dots, w_p)$ is a path in T_j and (because of property γ of T_j) is a B -chain; this fact and hypothesis $u_{i+1}vw$ imply that $(w_1, \dots, w_{h-1}, w_h, w_{h+1}, \dots, w_p)$ is itself a B -chain (due to Lemma 3.iv), hence $w_1w_hw_p$, i.e., xvy . The final step of the argument is the same as for case I. There is one last rule (the ninth), which is applied in

– Case V.5. The rule prescribes identifying a path $(u'_1, u'_2, \dots, u'_{m'})$ inside tree T_j so that $u'_1 = u_{i+1}$, $u'_2 = w$, and $u'_{m'}$ is a leaf. If $m' = 2$ (i.e., w itself is a leaf), then put $X_{j+1} = X_j \cup \{v\}$ and $A_{j+1} = A_j \cup \{\{w, v\}\}$, and it is seen that both properties β and γ are preserved in passing from T_j to $T_{j+1} = (X_{j+1}, A_{j+1})$. Otherwise, take path $(u'_1, \dots, u'_{m'})$ as a new reference for inserting point v inside tree T_j , by distinguishing the same cases and applying the same rules as used when comparing v with $(u_1, \dots, u_m)$. This is a recursive rule because, when comparing v with $(u'_1, \dots, u'_{m'})$, we may find that case V.5 occurs, which would imply that the rule must be applied once again when comparing v with some other path $(u''_1, \dots, u''_{m''})$ in the tree. Note, however, that paths $(u_1, \dots, u_m)$, $(u'_1, \dots, u'_{m'})$, $(u''_1, \dots, u''_{m''})$, ... possibly considered in successive applications of the rule, are inside smaller and smaller branches of tree T_j , which ensures that sooner or later the process reaches its end, i.e., point v comes to be inserted inside tree T_j so that properties β and γ are preserved. $\square$

Any tree-adjacency *uniquely* determines a tree-betweenness (Definition 2), and the converse is also true. This is because, if A and E are tree-adjacencies on the same domain X so that $A \neq E$, then there is some line $\{u, w\} \in A \setminus E$, which implies that, for any point v in the path joining u with w within tree (X, E) , it must be $(u, v, w) \in B_E \setminus B_A$, hence $B_A \neq B_E$. The constructive method used in proving Theorem 1 is suggestive of an algorithm for recovering a tree-adjacency based on the

associated tree-betweenness – i.e., determining a tree-adjacency based on a ternary relation satisfying conditions (2) and (4)–(7). It appears, from the proof, that such an algorithm would be a “backtrack-free” procedure, i.e., in applying it, there can be no step at which one must modify the result of some previous step in order to continue the process to its end. This is a good property, and implies that the number of main steps in the process will not exceed the number of points in domain X . From the proof, it also appears that the order in which the points in X may be considered in constructing the tree is virtually free: at each step along the process, any of the points not yet located in the current partial tree may be chosen to be inserted consistently in the tree.

In view of the correspondences noticed at the beginning of this section (second paragraph), Theorem 1 implies the following solution to the full representation problem.

COROLLARY 1. *Let X and Y be two sets (with $|X| \geq 3$), $\mathcal{X}$ a family of subsets of Y related by one-to-one correspondence to the elements of X , and $B_{\mathcal{X}}$ the intersection-betweenness it induces on X (Definition 1). Then there is a tree (X, A) such that $B_A = B_{\mathcal{X}}$ if and only if conditions (8), (9), and (10) are true of all distinct members of $\mathcal{X}$.*

Based on the proof of the theorem, we may see that, if conditional completeness (7) is replaced by absolute completeness:

$$\text{for all } \{u, v, w\} \subseteq X, uvw \text{ or } vwu \text{ or } wuv, \quad (13)$$

then a system (2),(4)–(6),(13) is obtained which characterises *chain-betweenness*. We may also see that the logical conjunction of (8) and the intersection-betweenness version of (13) equals the following condition, referred to all $\{U, V, W\} \subseteq \mathcal{X}$:

$$U \cap V^c \cap W = \emptyset \text{ if and only if } (U^c \cap V \cap W \neq \emptyset \text{ and } U \cap V \cap W^c \neq \emptyset). \quad (14)$$

All this implies the following result concerning *chain representation* of intersection-betweennesses.

COROLLARY 2. *In the general conditions of Corollary 1, there is a chain (X, A) so that $B_A = B_{\mathcal{X}}$ if and only if conditions (9) and (14) are true of all distinct members of family $\mathcal{X}$.*

3. CONDITIONS FOR PARTIAL TREE REPRESENTATION

In this section we discuss *partial* tree representation of intersection-betweenness, which is the second of the two problems defined in the Introduction. The information I present descends from noticing certain similarities between this representation problem and two special concepts discussed in separate branches of modern applied mathematics. One is the concept of a “join tree”, as defined in the theory of relational databases, which proves to be related to our problem in a direct way. The other is the concept of a “tree hypergraph”, as defined in the theory of combinatorial data analysis, which turns out to be related in an indirect (dual) way. By

highlighting basic similarities of both concepts with the partial tree representation problem, we come at the same time to see the connection between both concepts themselves, which is a step toward verifying correspondences between theoretical results regarding them³.

To permit of a “join tree” is a possible property of the “scheme” of a “relational database”, this scheme being a family $\mathcal{X}$ of subsets of a basic set Y (Y is the set of “attributes” involved in the database). The “complete intersection graph” for the database is the complete graph having $\mathcal{X}$ as the set of vertices and every edge $\{U, V\}$ (for $U \neq V \in \mathcal{X}$) labelled by all elements in $U \cap V$. For any $y \in Y$, any path in such a graph is called a “ y -pat” if y is among the labels of *every* edge in the path (cf. [Maier, 1983, §13.2.2]).

DEFINITION 4. *Let $\mathcal{X}$ be the scheme of a relational database on set of attributes Y , and A an adjacency relation on $\mathcal{X}$. Graph $(\mathcal{X}, A)$ is a join tree if it is connected and acyclic, and for all $U \neq V \in \mathcal{X}$, the path joining U and V in it is a y -path, for all $y \in U \cap V$.*

A database the scheme of which permits of a join tree is said to be “acyclic” (of course, arbitrary databases may fail to have this property). The importance of the concept is due to the fact that, when a database is acyclic in this sense, then some theoretical connections and computational procedures on the database itself become greatly simplified – e.g., in acyclic databases, “pairwise consistency” and “global consistency” are equivalent (cf. [Maier, 1983, §13.1.4]). Acyclicity has been subjected to mathematical scrutiny, and several characterisations of it have been discovered ([Beer, Fagin, Maier, Yannakakis, 1983] specify and prove eleven characterisations). Later on in this section we discuss one of these characterisations, that referring to the so-called “GYO algorithm”.

A general and obvious correspondence holds true between the acyclicity condition for families of sets (which may or not be the schemes of relational databases) and the solvability of the partial tree representation problem for intersection-betweennesses.

THEOREM 2. *Let R be a relation between sets X and Y such that $uR \neq vR$ for all $u \neq v \in X$ (so that elements of X and sets in $\mathcal{X} = \{xR : x \in X\}$ are linked by a natural one-to-one correspondence). Let A be a connected and acyclic adjacency relation on set X (or, equivalently, on family $\mathcal{X}$). Then (X, A) is a solution to the partial tree representation problem for the intersection-betweenness $B_{\mathcal{X}}$ induced by $\mathcal{X}$ on X if and only if $(\mathcal{X}, A)$ is a join tree.*

Proof. To prove part “if”, let us presume $(\mathcal{X}, A)$ be a join tree, and consider any triple $\{u, v, w\} \subseteq X$ such that $(u, v, w) \in B_A$, which means that $V = vR$ is in the path joining $U = uR$ and $W = wR$ within tree $(\mathcal{X}, A)$. That path is a y -path for all $y \in U \cap W$ (as $(\mathcal{X}, A)$ is a join tree), which implies $U \cap W \subseteq V$, i.e. $(u, v, w) \in B_{\mathcal{X}}$ (by Definition 1). By generalizing over triple $\{u, v, w\}$, we conclude $B_A \subseteq B_{\mathcal{X}}$, which

³In the original version of this paper, Section 3 only contained references to “join trees” and related results from the theory of relational databases. I introduced references to “tree hypergraphs” in the revised version, following a suggestion by one anonymous referee, who pointed out significant aspects of the examples in Section 4. I express my gratitude for this revealing observation.

means that tree (X, A) is a solution to the partial tree representation problem. Verifying part “only if” is equally easy. $\square$

This simple result has important consequences on our discussion. It implies that any of the known characterisations of database acyclicity – more precisely, any characterisation that refers only to the scheme, not other components of a database – may be interpreted and applied as a system of conditions singly necessary and jointly sufficient for the solvability of a partial tree representation problem (of an intersection-betweenness). It also implies that any procedure devised for the construction of a join tree may also be used for the construction of a tree partially representing an intersection-betweenness (when the betweenness allows of such a tree). We illustrate this possibility in Section 4, using the “GYO algorithm” as a test of solvability, and a related procedure for actually constructing a representing tree.

Let X be a set of objects and $\mathcal{Y}$ a family of its subsets – alternative names: $(X, \mathcal{Y})$ is a “hypergraph”, elements of X are the “vertices”, elements of $\mathcal{Y}$ are the “hyperedges”. Presume that the sets forming family $\mathcal{Y}$ have been empirically determined as answers to a classification task: for each $Z \in \mathcal{Y}$, all objects grouped into Z have been judged as mutually similar in some respect. If we decide to represent the empirical structure $(X, \mathcal{Y})$ by a tree (X, A) , then a natural requirement would be that each set $Z \in \mathcal{Y}$ specifies a sub-tree: in that case we would obtain that the mutual similarity between objects grouped into Z would be expressed as the connectedness of the corresponding set of vertices in the tree. This idea led Flament (1978) to introduce the concept of a “tree hypergraph” (“hypergraphe arboré”).

DEFINITION 5. *Let $(X, \mathcal{Y})$ be a hypergraph and (X, A) a tree, both defined on X as the set of vertices. The hypergraph is rigid on the tree if each of its hyperedges is a connected set of vertices in the tree – i.e., for all $u, v, w \in X$ and $Z \in \mathcal{Y}$, if $u \in Z$, $w \in Z$, and v is in the path joining u and w within the tree, then also $v \in Z$.*

Any hypergraph which allows of such a tree representation, i.e., which is rigid on some tree definable on its set of vertices, is called a “tree hypergraph”. Several characterisations of the concept are available, i.e., conditions ensuring that a family of sets can be represented in the stated form (cf. [Flament, 1978; Leclerc, 1987, §3.2]).

A regular connection holds true between solvability of our partial tree representation problem and the condition defining tree hypergraphs. This is a kind of “dual” connection, as the two conditions distinctly refer to families of sets determined on the one and the other side of the basic binary relation.

THEOREM 3. *Let R be a relation between sets X and Y , $\mathcal{X} = \{xR : x \in X\}$ and $\mathcal{Y} = \{Ry : y \in Y\}$ the families of subsets of Y and, respectively, of X it determines, and $B_{\mathcal{X}}$ the intersection-betweenness induced by $\mathcal{X}$ (Definition 1). Let A be a connected and acyclic adjacency relation on X . Then tree (X, A) is a partial tree representation of betweenness $B_{\mathcal{X}}$ if and only if family $\mathcal{Y}$ is rigid on it.*

Proof. – Part “if”: Let us presume that $\mathcal{Y}$ is rigid on (X, A) , and suppose – to reach a contradiction – that (X, A) does not partially represent betweenness $B_{\mathcal{X}}$.

This negative hypothesis means that $(u, v, w) \in B_A$ but $(u, v, w) \notin B_{\mathcal{X}}$ for some $u, v, w \in X$, i.e., v is in the path joining u and w within tree (X, A) , but there exists some element z in $U \cap V^c \cap W$ (with $U = uR$, $V = vR$ and $W = wR$, which are members of $\mathcal{X}$). This would imply $\{u, w\} \subseteq Z$ but $v \notin Z$ (with $Z = Rz \in \mathcal{Y}$), which is impossible, because v is in the path joining u and w within tree (X, A) , and family $\mathcal{Y}$ is presumed rigid on this tree.

– Part “only if”: Conversely, let us presume that tree (X, A) partially represents intersection-betweenness $B_{\mathcal{X}}$ (i.e., $B_A \subseteq B_{\mathcal{X}}$), consider any $Z = Rz$ member of family $\mathcal{Y}$, and refer to any triple $\{u, v, w\} \subseteq X$ so that $u, w \in Z$ and v is in the path joining u and w within the tree (i.e., $(u, v, w) \in B_A$). In these conditions, $(u, v, w) \in B_{\mathcal{X}}$, i.e., $U \cap W \subseteq V$ (with $U = uR$, $V = vR$ and $W = wR$), hence $z \in V$ (because $z \in U \cap W$), so that $v \in Z$. By generalizing over triple $\{u, v, w\}$, we reach the conclusion that set Z specifies a sub-tree of tree (X, A) . This is true of every $Z \in \mathcal{Y}$, hence family $\mathcal{Y}$ is rigid on the presumed tree. $\square$

This result shows that, by simple shift of perspective (from one to the other side of the basic binary relation), any property relating to tree hypergraphs (in the sense of [Flament, 1978]) can be translated into a property referring to intersection-betweennesses and their partial tree representation, and vice-versa. The same is true of properties concerning tree hypergraphs and join trees, due to the direct connection linking our definition of a partial tree representation to this last concept (Theorem 2). All this leads us to surmise possible fruitful exchanges in the mathematical study of both concepts – the one within the theory of relational databases, the other within the theory of combinatorial data analysis – and to expect correspondences between already established facts concerning them. The following are two examples: the characterisation of tree hypergraphs in terms of Helly’s property and triangulated graphs (cf. [Flament, 1978, Th. 5]) corresponds to the characterisation of acyclic databases in terms of chordal graphs and conformal hypergraphs (cf. [Beeri *et al.*, 1983, Th. 3.4.3]), and the characterisation of tree hypergraphs in terms of maximum trees on their traces (cf. [Flament, 1978, Th. 13]) corresponds to the characterisation of acyclic databases in terms of maximum weight spanning trees (cf. [Bernstein, Goodman, 1981, §5]).

From this point on in this Section I focus on one particular characterisation of acyclicity of a database, which – thanks to Theorem 2 – may serve as a test for deciding whether the intersection-betweenness induced by a family of sets can or cannot be partially represented by a tree. The characterisation is centred on the so-called “GYO algorithm”, independently proposed by Graham [1979] and Yu, Ozsoyoglu [1979] (the initials of the three surnames form the name of the algorithm). I prefer to discuss this characterisation for several reasons: it only involves consideration of the scheme of a database, i.e., a family of sets (other characterisations of acyclicity also involve reference to other components of a database, which have no counterpart in our representation problem); it is comparatively simple (among the characterisations sharing the property just mentioned); in practice, it takes the form of a sequence of elementary operations, governed by few basic rules; in proving it, the profile of a procedure will emerge for actually constructing a representing tree (when the conditions for the existence of such a tree are satisfied). The fact that the “success” of the GYO algorithm with a family of sets characterises the

acyclicity of the family, is a well-known result of the theory of relational databases. Nevertheless, I think it fit for my study not only to recall this fact (Theorem 4), but also to explicitly prove it, for two reasons. One is to present a self-contained proof of that characterisation, in the terms and within the frame of this paper (proofs readily available of that result are given as parts of larger argumentations, which involve several other characterisations of “database acyclicity”, these being discussed through chains of implications (cf. [Beer *et al.*, 1983, §6; Maier, 1983, §13.3.3; Abiteboul, Hull, Vianu, 1995, §6.4]). The other reason is one mentioned just above: in proving the chosen characterisation – more precisely, in developing the “if” part of the proof – the essential steps will emerge of a procedure for the construction of a tree partially representing an intersection-betweenness (whenever such a tree does exist). The procedure is outlined after completing the proof of Theorem 4, and illustrated in Section 4 (second example).

The *input* of the *GYO algorithm* may be any (ordered) family of sets $\mathcal{X} = (U_1, \dots, U_m)$. Starting from it, the algorithm produces a sequence of families:

$$\begin{aligned} \mathcal{X}_1 &= (U_{1,1}, \dots, U_{1,m}) = \mathcal{X} \\ &\dots\dots\dots \\ \mathcal{X}_j &= (U_{j,1}, \dots, U_{j,m}) \\ \mathcal{X}_{j+1} &= (U_{j+1,1}, \dots, U_{j+1,m}) \\ &\dots\dots\dots \end{aligned}$$

which is component-wise monotone (i.e., $U_{j,i} \supseteq U_{j+1,i}$ for all $j \in \{1, 2, \dots\}$ and $i \in \{1, \dots, m\}$). The passage from any family $\mathcal{X}_j$ to its successor $\mathcal{X}_{j+1}$ may consist in one of two operations, called “reduction” and “depletion”.

The *reduction operation*: A component $U_{j,i}$ (if there is any) is identified, so that this set-theoretic difference is non-empty:

$$U_{j,i} \setminus (U_{j,1} \cup \dots \cup U_{j,i-1} \cup U_{j,i+1} \cup \dots \cup U_{j,m});$$

one element x is chosen in this difference; family $\mathcal{X}_{j+1}$ is formed by setting:

$$U_{j+1,i} = U_{j,i} \setminus \{x\}, \quad U_{j+1,h} = U_{j,h} \text{ for all } h \neq i.$$

The *depletion operation*: A component $U_{j,i}$ (if there is any) is identified, so that:

$$\emptyset \neq U_{j,i} \subseteq U_{j,k} \text{ for some } k \neq i;$$

family $\mathcal{X}_{j+1}$ is formed by setting:

$$U_{j+1,i} = \emptyset, \quad U_{j+1,h} = U_{j,h} \text{ for all } h \neq i.$$

The process terminates when a family $\mathcal{X}_n = (U_{n,1}, \dots, U_{n,m})$ is obtained on which neither a reduction nor a depletion can be applied (this means that $U_{n,i} \subseteq \bigcup (\mathcal{X}_n \setminus \{U_{n,i}\})$ for all $1 \leq i \leq m$, and $(U_{n,i} = \emptyset \text{ or not}(U_{n,i} \subseteq U_{n,k}))$ for all $1 \leq i \neq k \leq m$). The resulting family $\mathcal{X}_n = (U_{n,1}, \dots, U_{n,m})$ is the *output* of the algorithm. The algorithm is said to *succeed* with input $\mathcal{X}$ when $\mathcal{X}_n = (\emptyset, \dots, \emptyset)$.

Note that, starting from one family of sets, the GYO algorithm may produce different sequences of families in separate complete implementations. This is so because, at any step in the process, there may be different components in the current family and different elements in the chosen component on which to perform a reduction, or different components on which to perform a depletion. It is shown, however, that the last family in the process – the output of the algorithm – is independent of the way in which the algorithm is implemented (in technical terms, the GYO algorithm has the “Church-Rosser property”; cf. [Beeri *et al.*, 1983, p. 485]. In particular, the fact that the algorithm succeeds or not with a given family of sets is an invariant property of that family.

Precisely the “success” of the GYO algorithm with any given family of sets characterises the “acyclicity” of that family, hence the solvability of the representation problem we are discussing.

THEOREM 4. *For any family of sets $\mathcal{X}$, there is an adjacency A so that $(\mathcal{X}, A)$ is a tree partially representing the intersection-betweenness inherent in $\mathcal{X}$, if and only if the GYO algorithm succeeds with $\mathcal{X}$.*

Proof. – Part “only if”: Presume that on a family of sets $\mathcal{X} = (U_1, \dots, U_m)$ an adjacency A can be defined so that $(\mathcal{X}, A)$ is a tree and for all $i, i', i'' \in \{1, \dots, m\}$, if $U_{i'}$ is in the path from U_i to $U_{i''}$ within the tree, then $U_i \cap U_{i''} \subseteq U_{i'}$. Choose any element $r \in M = \{1, \dots, m\}$ and define a permutation $(g(1), \dots, g(m))$ of set M so that $g(1) = r$ and for all $g(i) \neq g(j) \in M \setminus \{r\}$, if $U_{g(i)}$ is in the path from $U_{g(1)}$ to $U_{g(j)}$ within tree $(\mathcal{X}, A)$, then $i < j$. Also define a function f so that $U_{g(f(j))}$ is the immediate precursor of $U_{g(j)}$ in the path from $U_{g(1)}$ to $U_{g(j)}$ within the tree, for all $j \in M \setminus \{1\}$. If $i < j$, then the path joining $U_{g(i)}$ with $U_{g(j)}$ within the tree passes through $U_{g(f(j))}$, which implies $U_{g(i)} \cap U_{g(j)} \subseteq U_{g(f(j))}$. This being true of all $i < j$, the following condition is proven (family $\mathcal{X}$ satisfying this condition is said to have the “running intersection property”):

$$U_{g(j)} \cap (U_{g(1)} \cup U_{g(2)} \cup \dots \cup U_{g(j-1)}) \subseteq U_{g(f(j))}, \text{ for all } j \in M \setminus \{1\}. \quad (15)$$

Now consider component $U_{g(m)}$ of family $\mathcal{X}$, and split it into two parts (one of which may be empty):

$$\begin{aligned} U'_{g(m)} &= U_{g(m)} \setminus (U_{g(1)} \cup U_{g(2)} \cup \dots \cup U_{g(m-1)}) \\ U''_{g(m)} &= U_{g(m)} \cap (U_{g(1)} \cup U_{g(2)} \cup \dots \cup U_{g(m-1)}). \end{aligned}$$

Through a suitable run of GYO reduction operations, all elements of $U'_{g(m)}$ can be cancelled, so that at the end of the run we have $U''_{g(m)}$ in place of $U_{g(m)}$. But $U''_{g(m)} \subseteq U_{g(f(m))}$ due to (15), so that, through a GYO depletion operation, component $U''_{g(m)}$ can be replaced by empty set $\emptyset$. This argument can be in turn applied to components $U_{g(m-1)}, U_{g(m-2)}, \dots$, and shows that each set in family $\mathcal{X}$ can be substituted by empty set $\emptyset$ through suitable applications of the GYO operations. Thus, the GYO algorithm succeeds with family $\mathcal{X}$.

– Part “if”: Let $\mathcal{X} = (U_1, \dots, U_m)$ be a family of (non-empty) sets with which the GYO algorithm does succeed. Then consider the sequence of families produced by any GYO process, i.e., a complete implementation of the algorithm:

$$\mathcal{X}_1 = (U_{1,1}, \dots, U_{1,m}), \mathcal{X}_2 = (U_{2,1}, \dots, U_{2,m}), \dots, \mathcal{X}_n = (U_{n,1}, \dots, U_{n,m}).$$

Thus:

$$\emptyset \neq U_i = U_{1,i} \supseteq U_{2,i} \supseteq \dots \supseteq U_{n-1,i} \supseteq U_{n,i} = \emptyset, \text{ for all } i \in M = \{1, \dots, m\}.$$

Then, for each $i \in M$ put:

$$f(i) = \min\{j \in \{1, \dots, n\} : U_{j,i} = \emptyset\},$$

which implies $U_{f(i)-1,i} \neq \emptyset = U_{f(i),i}$. This means that component $U_{f(i)-1,i}$ is the specific subject of the operation applied in passing from $\mathcal{X}_{f(i)-1}$ to $\mathcal{X}_{f(i)}$. If the operation is a reduction and $f(i) < n$, then choose one element in $M \setminus \{i\}$ and denote it by $o(i)$. If, instead, the operation is a depletion, then there must be some element $h \in M \setminus \{i\}$ so that $U_{f(i)-1,i} \subseteq U_{f(i)-1,h}$; hence, choose one such element and denote it by $o(i)$. In this way, a set of ordered pairs is formed:

$$O^\circ = \{(1, o(1)), \dots, (s-1, o(s-1)), (s+1, o(s+1)), (m, o(m))\},$$

which is a function from $M \setminus \{s\}$ to M , term s being such that $f(s) = n$ (i.e., s is the index of the last component of family $\mathcal{X}$ to be emptied in the process). Clearly:

$$f(u) < f(v) \text{ for all } (u, v) \in O^\circ \quad (16)$$

i.e., if $v = o(u)$, then the v -th component of $\mathcal{X}$ becomes empty one or more steps after the u -th component. Based on these data, let us define the following (undirected) graph:

$$T' = (M, O) = (\{1, \dots, m\}, \{\{1, o(1)\}, \dots, \{s-1, o(s-1)\}, \{s+1, o(s+1)\}, \{m, o(m)\}\}).$$

The graph is acyclic. Presume – to reach a contradiction – that $(u_1, \dots, u_t)$ is a cycle in the graph. Then, for no $k \in \{1, \dots, t\}$ it may be $(u_k, u_{k-1}), (u_k, u_{k+1}) \in O^\circ$ (because O° is a function), nor for any $k \in \{1, \dots, t\}$ it may be $(u_{k-1}, u_k), (u_{k+1}, u_k) \in O^\circ$ (since this would imply $(u_h, u_{h-1}), (u_h, u_{h+1}) \in O^\circ$ for some $h \neq k$, which is impossible). Further, it cannot be $(u_1, u_2), \dots, (u_{t-1}, u_t), (u_t, u_1) \in O^\circ$, since in this case $s \notin \{u_1, \dots, u_t\}$ (as no ordered pair in O° has s as its first component), and $f(u_1) < f(u_2) < \dots < f(u_t) < f(u_1)$ (according to (16)), which is obviously contradictory. For similar reasons, case $(u_1, u_t), (u_t, u_{t-1}), \dots, (u_2, u_1) \in O^\circ$ is also impossible. Thus, graph T' cannot contain any cycle. This means that $T' = (M, O)$ is a forest. The forest can be transformed into a tree $T = (M, A) = (M, O \cup E)$ by adding a suitable set of lines E – if T' is already a tree, i.e., it is a connected graph, then $E = \emptyset$. What remains to prove is that tree T partially represents the intersection-betweenness inherent in the family of sets $\mathcal{X} = (U_1, \dots, U_m)$. We prove this by contradiction, starting from the hypothesis that T does not have the alleged capacity towards $\mathcal{X}$, which means that there is some path $(p_1, \dots, p_h)$ in T so that:

$$\emptyset \neq U_{p_1} \cap U_{p_h} \supset U_{p_1} \cap \dots \cap U_{p_h}. \quad (17)$$

Let $(q_1, \dots, q_k)$ be a path in the tree which satisfies (17) and is such that $\min(f(q_1), f(q_k)) \geq \min(f(p_1), f(p_h))$ for all paths $(p_1, \dots, p_h)$ satisfying (17); we may presume $f(q_1) < f(q_k)$. Because $U_{q_1} \cap U_{q_k} \neq \emptyset$ and $U_{f(q_1)-1, q_1} \neq \emptyset \neq U_{f(q_1)-1, q_k}$,

no point in $U_{q_1} \cap U_{q_k}$ may have been deleted through a reduction operation in the first $f(q_1) - 1$ steps of the GYO process, hence:

$$U_{f(q_1)-1, q_1} \cap U_{f(q_1)-1, q_k} = U_{q_1} \cap U_{q_k}. \quad (18)$$

This implies that the $f(q_1)$ -th step in the GYO process must be a depletion operation, acting upon $U_{f(q_1)-1, q_1}$. Thus, $o(q_1) \in M \setminus \{q_1\}$ is such that $U_{f(q_1)-1, q_1} \subseteq U_{f(q_1)-1, o(q_1)}$, and this implies:

$$U_{f(q_1)-1, q_1} \cap U_{f(q_1)-1, q_k} \subseteq U_{f(q_1)-1, o(q_1)} \cap U_{f(q_1)-1, q_k} \subseteq U_{o(q_1)} \cap U_{q_k}. \quad (19)$$

Now, $o(q_1) \neq q_k$ (because $(q_1, o(q_1))$ is a path in T and does not satisfy (17)) and $\min(f(q_1), f(q_k)) < \min(f(o(q_1)), f(q_k))$ (due to (16) and hypothesis $f(q_1) < f(q_k)$), so that, if $(r_1, \dots, r_l)$ is the path in T joining $r_1 = o(q_1)$ with $r_l = q_k$, then, due to the choice of path $(q_1, \dots, q_k)$, we infer that $(r_1, \dots, r_l)$ does *not* satisfy (17), which means (because $U_{r_1} \cap U_{r_l} \supseteq U_{q_1} \cap U_{q_k} \neq \emptyset$):

$$U_{r_1} \cap U_{r_l} = U_{r_1} \cap \dots \cap U_{r_l}. \quad (20)$$

Now, let us compare paths $(q_1, \dots, q_k)$ and $(r_1, \dots, r_l)$. Considering that T is a tree, $\{q_1, r_1\}$ (i.e., $\{q_1, o(q_1)\}$) is a line in the tree, and $q_k = r_l$, we infer that:

$$\text{either } (q_2, \dots, q_k) = (r_1, r_2, \dots, r_l) \text{ or } (q_1, q_2, \dots, q_k) = (r_2, \dots, r_l).$$

In the former case, by combining (18), (19) and (20), we derive $U_{q_1} \cap U_{q_k} \subseteq U_{q_2} \cap \dots \cap U_{q_k}$, which contradicts (17) as referred to $(q_1, \dots, q_k)$. In the latter case, by combining (18), (19) and (17) as referred to $(q_1, \dots, q_k)$, we derive $U_{r_1} \cap U_{r_l} \supset U_{r_2} \cap \dots \cap U_{r_l}$, which contradicts (20). $\square$

Part “if” of this proof shows that a full implementation of the GYO algorithm on a family of sets can serve not only to test the solvability of the partial tree representation problem for that family, but also to construct a solution to the problem (when the algorithm succeeds with the family). The constructive procedure may be outlined as follows.

- (i) Start and continue to its end an implementation of the GYO algorithm on the assigned family of sets $\mathcal{X} = (U_1, \dots, U_m)$.
- (ii) For any j -th step of the process at which some component U_i of $\mathcal{X}$ becomes empty, sign a line joining U_i to some other component $U_{o(i)}$ which is chosen among those such that $U_{j-1, h} \neq \emptyset$ or $U_{j-1, h} \supseteq U_{j-1, i}$, depending on whether the j -th operation in the GYO process is a reduction or a depletion.
- (iii) The set of lines O thus determined is an acyclic adjacency on $\mathcal{X}$. If O is connected, then $(\mathcal{X}, A) = (\mathcal{X}, O)$ is a solution to the partial representation problem. If O is not connected, then add a suitable set of lines E connecting the trees in forest $(\mathcal{X}, O)$, and the resulting tree $(\mathcal{X}, A) = (\mathcal{X}, O \cup E)$ is a solution to the partial representation problem.

Note that the procedure is backtrack-free: there can be no step in the process at which, in order to continue the process to its end, we must undo something already done in previous steps. Also note that there are passages in the procedure at which

a step may be made in different ways: for one and the same family $\mathcal{X}$ there are, in general, several ways of fully implementing the GYO algorithm; when a component U_i becomes empty due to a reduction or depletion operation, then there may be several components $U_h \neq U_i$ so that $U_{j-1,h} \neq \emptyset$ or $U_{j-1,h} \supseteq U_{j-1,i}$, i.e., several components which we may choose to link up with U_i ; lastly, if the resulting set of lines O is not connected, then we may differently choose the additional set of lines E so that $(\mathcal{X}, O \cup E)$ is a tree. All these aspects point to the fact that, if a problem of partial tree representation is solvable, then it generally admits *several* legitimate solutions. This contrasts with the uniqueness of the solution to a full tree representation problem (when it is solvable), mentioned after the proof of Theorem 1. Actually, for full tree representation a tree $(\mathcal{X}, A)$ is sought for so that $B_A = B_{\mathcal{X}}$, whereas for partial tree representation a tree is sought for so that $B_A \subseteq B_{\mathcal{X}}$: intuitively, the larger the intersection-betweenness $B_{\mathcal{X}}$ (compared with tree-betweennesses), the looser the partial representation offered by tree-adjacency A , and the larger the number of tree-adjacencies satisfying the required condition.

4. EXAMPLES

Let $X = \{a, b, \dots, j\}$ be a collection of ten seaside resorts, and suppose that a tour operator in a city is interested in discovering the internal structure of the collection, based on “semantic proximities” between resorts, as they are perceived by potential tourists. Imagine that, to this aim, the list of resorts is submitted to a sample of people $Y = \{y_1, \dots, y_{20}\}$ from the same city. All participants are asked to mark those resorts (at least two) in which they would like to spend their summer holidays. Let the data from such a hypothetical experiment be those in Table 2.

The information in Table 2 may equivalently be expressed as a family $\mathcal{X} = \{A, B, \dots, J\}$ of subsets of set Y . E.g., $A = \{y_1, y_2, y_3, y_4, y_5, y_{14}, y_{15}, y_{16}, y_{19}\}$ is the set having the first column of Table 2 as its indicator function, i.e., the set of participants in the experiment who give a positive answer for resort a . Through a number of comparisons between the sets in family $\mathcal{X}$, the conclusion can be reached that the family fully satisfies conditions (8)–(10). E.g., $(A + B) \cap C = \{y_6, y_{10}, y_{12}, y_{18}, y_{20}\} \neq \emptyset$, as required by (8); $A \cap B^c \cap C = \emptyset$ and $(A \cap C^c \cap D = \{y_1, y_2, y_{14}, y_{15}\} \neq \emptyset$ and $B \cap C^c \cap D = \{y_2, y_{11}, y_{15}\} \neq \emptyset$), which complies with (9); $A^c \cap C \cap E = \{y_{12}, y_{18}\}$, $A \cap C^c \cap E = \{y_{15}\}$, $A \cap C \cap E^c = \{y_{16}, y_{19}\} \neq \emptyset$ but $A \cap B^c \cap C \cap E = \emptyset$, which complies with (10); and so on. As the three conditions are never falsified within family $\mathcal{X}$, by virtue of Corollary 1 we infer that the intersection-betweenness $B_{\mathcal{X}}$ induced by $\mathcal{X}$ on X (which is described in Table 3) can be fully represented by a tree.

The next step is to derive the complementary relation $N_{\mathcal{X}}$ of betweenness $B_{\mathcal{X}}$ and the associated median function $c_{\mathcal{X}}$: these are needed in order to apply rules of Table 1 in constructing the tree representation of the given intersection-betweenness. Complementary relation $N_{\mathcal{X}}$ is the set of unordered triples listed in Table 4, and median function $c_{\mathcal{X}}$ associates points b , d , and h with each of the triples in the first, second, and third parts of the table, respectively.

Now we are in a position to apply the procedure outlined in the proof of Theorem 1 for constructing the tree fully representing the obtained qualitative proximities. A complete implementation of the procedure is shown in Figure 2. The graph in

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>	<i>j</i>
y_1	1	0	0	1	0	0	0	0	0	1
y_2	1	1	0	1	0	0	0	0	1	1
y_3	1	1	1	1	1	0	0	0	1	1
y_4	1	1	1	1	1	0	1	1	1	1
y_5	1	1	1	1	1	1	1	1	1	0
y_6	0	1	1	0	0	1	1	1	1	0
y_7	0	0	1	0	0	1	1	1	0	0
y_8	0	0	0	0	0	1	1	1	0	0
y_9	0	1	0	0	1	0	0	0	1	0
y_{10}	0	1	1	1	0	1	0	1	0	1
y_{11}	0	1	0	1	1	0	0	0	1	1
y_{12}	0	1	1	0	1	1	0	1	1	0
y_{13}	0	0	0	0	1	0	0	0	1	0
y_{14}	1	0	0	1	0	0	0	0	0	1
y_{15}	1	1	0	1	1	0	0	0	1	0
y_{16}	1	1	1	1	0	1	0	1	0	0
y_{17}	0	0	0	0	0	1	1	1	0	0
y_{18}	0	1	1	0	1	0	1	1	1	0
y_{19}	1	1	1	1	0	0	1	1	0	0
y_{20}	0	1	1	1	0	0	1	1	0	1

Table 2. Data from a hypothetical choice experiment on a list of 10 seaside resorts (column labels) judged by 20 participants (row labels). Each row records the answers by one participant (number 1 for a positive answer).

the last step (T_7) is the full tree representation of the presumed experimental data, as may be seen by verifying that the betweenness inherent in the graph is precisely the ternary relation of Table 3. The resulting tree reveals the internal organisation of the set of seaside resorts, as they are perceived by the sample of participants in the experiment.

As another example, let us presume that $X = \{a, b, \dots, l\}$ is a set of twelve “attributes” having psychological significance (habits, personality traits, crucial past experiences, etc.), that a sample of persons $Y = \{y_1, \dots, y_{15}\}$ have been examined relative to such attributes, and that the collected data are those of Table 5. In this hypothetical survey, all attributes are treated as dichotomous variables, so that, for each attribute and each person, the empirical information is whether the person has or has not the attribute (numbers 1 and 0 for positive and negative answers). We are interested in making explicit the internal organisation of the set of attributes, based on qualitative proximities derivable from the data, and presuming that the organisation of the set is acyclic in character, i.e., it may be represented as a tree.

The columns of Table 5 are the indicator functions of different subsets of set Y forming the family $\mathcal{X} = \{A, B, \dots, L\}$, with $A = \{y_1, y_2, y_5, y_7, y_8, y_{12}\}$, $B = \{y_2, y_3, y_6, y_7, y_8, y_9, y_{11}, y_{13}, y_{14}\}$, and so on. The intersection-betweenness $B_{\mathcal{X}}$ induced on set X by family $\mathcal{X}$ is described in Table 6. It turns out that this relation includes some violations of conditions (4)–(6) (which means that family $\mathcal{X}$ does not fully comply with requirements (8) and (9)). For example, both triples abg and agb are in relation $B_{\mathcal{X}}$, which contradicts condition (4); both abc and bcd , but not acd , are in $B_{\mathcal{X}}$, which contradicts condition (5); both acf and abc , but not bcf , are in $B_{\mathcal{X}}$, which contradicts condition (6). Thus, due to Theorem 1 (part “only

	<i>j</i>	<i>i</i>	<i>h</i>	<i>g</i>	<i>f</i>	<i>e</i>	<i>d</i>	<i>c</i>	<i>b</i>
<i>a</i>	<i>d</i>	<i>b, d</i>	<i>b, c, d</i>	<i>b, c, d, h</i>	<i>b, c, d, h</i>	<i>b, d, i</i>		<i>b, d</i>	<i>d</i>
<i>b</i>	<i>d</i>		<i>c</i>	<i>c, h</i>	<i>c, h</i>	<i>i</i>			
<i>c</i>	<i>b, d</i>	<i>b</i>		<i>h</i>	<i>h</i>	<i>b, i</i>	<i>b</i>		
<i>d</i>		<i>b</i>	<i>b, c</i>	<i>b, c, h</i>	<i>b, c, h</i>	<i>b, i</i>			
<i>e</i>	<i>b, d, i</i>		<i>b, c, i</i>	<i>b, c, h, i</i>	<i>b, c, h, i</i>				
<i>f</i>	<i>b, c, d, h</i>	<i>b, c, h</i>		<i>h</i>					
<i>g</i>	<i>b, c, d, h</i>	<i>b, c, h</i>							
<i>h</i>	<i>b, c, d</i>	<i>b, c</i>							
<i>i</i>	<i>b, d</i>								

Table 3. Intersection-betweenness $B_{\mathcal{X}}$ induced by columns of Table 2. E.g., entry d in the upper-left cell means that triples adj and jda are in relation $B_{\mathcal{X}}$ (because $A \cap J \subseteq D$). Similarly, entries b and d in the cell next on the right mean that triples abi , iba , adi , ida are in relation $B_{\mathcal{X}}$.

B -incomparable triples	median
$\{a, e, c\}, \{a, e, f\}, \{a, e, g\}, \{a, e, h\}, \{a, i, c\}, \{a, i, f\}, \{a, i, g\}, \{a, i, h\},$ $\{d, e, c\}, \{d, e, f\}, \{d, e, g\}, \{d, e, h\}, \{d, i, c\}, \{d, i, f\}, \{d, i, g\}, \{d, i, h\},$ $\{j, e, c\}, \{j, e, f\}, \{j, e, g\}, \{j, e, h\}, \{j, i, c\}, \{j, i, f\}, \{j, i, g\}, \{j, i, h\}$	b
$\{a, j, b\}, \{a, j, c\}, \{a, j, e\}, \{a, j, f\}, \{a, j, g\}, \{a, j, h\}, \{a, j, i\}$	d
$\{f, g, a\}, \{f, g, b\}, \{f, g, c\}, \{f, g, d\}, \{f, g, e\}, \{f, g, i\}, \{f, g, j\}$	h

Table 4. Complementary relation $N_{\mathcal{X}}$ and median function $c_{\mathcal{X}}$ associated with intersection-betweenness of Table 3.

if”), we must conclude that the qualitative proximity structure in this example has *no full* tree representation, i.e., there is no tree-betweenness equal to the obtained intersection-betweenness.

Given this negative result, we may then ask whether the data have a *partial* tree representation, i.e., a tree-betweenness which is (properly) included in the obtained intersection-betweenness. This possibility may be tested by applying the GYO algorithm to the family of sets $\mathcal{X}$, as suggested by Theorem 4. The main stages of a complete implementation of the algorithm are described in Table 7. The passage from the first to the second part of the table consists of seven depletions, denoted by δ (e.g., by writing $\delta(C; G)$, we mean that set C_1 can be depleted, because $C_1 \subseteq G_1$; similarly, by writing $\delta(D; E, H, K)$, we mean that set D_1 can be depleted, because $D_1 \subseteq E_1$, as well as $D_1 \subseteq H_1$ and $D_1 \subseteq K_1$). The passage from the second to the third part of the table consists of three reductions, denoted by ρ (e.g., by writing $\rho(G, y_{10})$, we mean that set G_8 can be reduced to $G_9 = G_8 \setminus \{y_{10}\}$,

FIGURE 2. Step-by-step construction of the tree fully representing the intersection-betweenness from the binary relation of Table 2. Each step is specified by declaring the maximal path chosen for reference, the new element considered for insertion, and the rule (among those in Table 1) to be applied in insertion.

as $y_{10} \in G_8 \setminus (A_8 \cup \dots \cup F_8 \cup H_8 \cup \dots \cup L_8)$.

We see that all sets in the last family of the process are identical to the empty set, i.e., the GYO algorithm succeeds with family $\mathcal{X}$, so that our question is answered in the positive: there is some tree partially representing the qualitative proximities of Table 5. To identify such a tree, let us apply the procedure stated after the proof of Theorem 4, by specifically referring to the GYO implementation in Table 7. The following set is a collection of ordered pairs which is consistent with that implementation:

$$O^\circ = \{(c, g), (d, h), (f, b), (i, e), (j, h), (k, h), (l, a), (a, e), (g, b), (h, e), (b, e)\}.$$

Specifically, pair (c, g) is accepted in the first step of the process (when component C_1 is depleted due to inclusion $C_1 \subseteq G_1$); pair (d, h) may be accepted in the second step (when component D_2 may be depleted by considering that $D_2 \subseteq H_2$); pair

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>	<i>j</i>	<i>k</i>	<i>l</i>
<i>y</i> ₁	1	0	0	1	1	0	0	1	1	1	1	1
<i>y</i> ₂	1	1	1	0	1	1	1	0	1	0	0	1
<i>y</i> ₃	0	1	1	1	1	1	1	1	0	1	1	0
<i>y</i> ₄	0	0	0	0	0	0	0	1	0	1	0	0
<i>y</i> ₅	1	0	0	0	1	0	0	1	1	1	0	1
<i>y</i> ₆	0	1	0	0	1	0	0	1	1	1	0	0
<i>y</i> ₇	1	1	0	0	1	0	1	1	0	0	0	0
<i>y</i> ₈	1	1	1	0	1	0	1	0	1	0	0	1
<i>y</i> ₉	0	1	1	0	0	0	1	0	0	0	0	0
<i>y</i> ₁₀	0	0	1	0	0	0	1	0	0	0	0	0
<i>y</i> ₁₁	0	1	0	0	1	1	0	1	0	0	1	0
<i>y</i> ₁₂	1	0	0	1	1	0	0	1	0	0	1	1
<i>y</i> ₁₃	0	1	1	0	1	1	1	1	0	1	0	0
<i>y</i> ₁₄	0	1	1	0	0	1	1	0	0	0	0	0
<i>y</i> ₁₅	0	0	0	0	0	0	0	1	0	1	0	0

Table 5. Hypothetical incidence matrix of a set of 12 attributes (column labels) on a sample of 15 persons (row labels).

(a, e) is accepted in the 11-th step (when component A_{11} is depleted due to inclusion $A_{11} \subseteq E_{11}$); and so on. It is seen that set O of unordered pairs corresponding to ordered pairs in O° is an acyclic and connected adjacency on domain X , i.e., graph (X, O) is a tree. Its diagram is shown in the left part of Figure 3. It is easily verified that the betweenness relation of this tree is included in the intersection-betweenness of Table 6, which confirms that the tree is a suitable (partial) representation of qualitative proximities from the presumed data.

Note that, based on the implementation in Table 6, or some other complete implementation of the GYO algorithm, we may determine other sets of ordered pairs, which specify alternative trees suitable for the same family of sets $\mathcal{X}$. One example is the following set of pairs, which yields the tree depicted on the right of Figure 3:

$$U^\circ = \{(a, e), (c, g), (d, e), (f, b), (i, e), (j, h), (k, e), (l, e), (g, b), (h, e), (b, e)\}.$$

This illustrates the general property that, if an intersection-betweenness has a partial (but not full) tree representation, then it has more than one such representation.

5. COMPARISONS AND CONCLUSIONS

As stated in the Introduction, the data lying at the root of the representation problems in this paper are typically a relation R between two sets X and Y . In set-theoretic terms, R is a subset of $X \times Y$, but there are alternative ways of coding the information it contains. For example, R may be specified through a Boolean matrix having columns and rows labelled by elements of X and Y , respectively, so that each cell (x, y) (for $x \in X$ and $y \in Y$) contains numbers 1 or 0, depending on

	<i>l</i>	<i>k</i>	<i>j</i>	<i>i</i>	<i>h</i>	<i>g</i>	<i>f</i>	<i>e</i>	<i>d</i>	<i>c</i>	<i>b</i>
<i>a</i>	<i>e</i>	<i>d, e, h, l</i>	<i>e, h, i, l</i>	<i>e, l</i>	<i>e</i>	<i>b, e</i>	<i>b, c, e, g, i, l</i>		<i>e, h, k, l</i>	<i>b, e, g, i, l</i>	<i>e, g</i>
<i>b</i>	<i>a, c, e, g, i</i>	<i>e, f, h</i>	<i>e, h</i>	<i>e</i>	<i>e</i>				<i>c, e, f, g, h, j, k</i>	<i>g</i>	
<i>c</i>	<i>a, b, e, g, i</i>	<i>b, d, e, f, g, h, j</i>	<i>b, e, f, g, h</i>	<i>a, b, e, g, l</i>	<i>b, e, f, g, j</i>		<i>b, g</i>	<i>b, g</i>	<i>b, e, f, g, h, j, k</i>		
<i>d</i>	<i>a, e, h, k</i>	<i>e, h</i>	<i>e, h, k</i>	<i>a, e, h, j, k, l</i>	<i>e, k</i>	<i>b, c, e, f, h, j, k</i>	<i>b, c, e, g, h, j, k</i>	<i>h, k</i>			
<i>e</i>	<i>a</i>	<i>h</i>	<i>h</i>			<i>b</i>	<i>b</i>				
<i>f</i>	<i>a, b, c, e, g, i</i>	<i>b, e, h</i>	<i>b, c, e, g, h</i>	<i>a, b, c, e, g, l</i>	<i>b, e</i>	<i>b, c</i>					
<i>g</i>	<i>a, b, c, e, i</i>	<i>b, c, d, e, f, h, j</i>	<i>b, c, e, f, h</i>	<i>a, b, c, e, l</i>	<i>b, e</i>						
<i>h</i>	<i>a, e</i>	<i>e</i>		<i>e, j</i>							
<i>i</i>	<i>a, e</i>	<i>a, d, e, h, j, l</i>	<i>e, h</i>								
<i>j</i>	<i>a, e, h, i</i>	<i>d, e, h</i>									
<i>k</i>	<i>a, d, e, h</i>										

 Table 6. Intersection-betweenness $B_{\mathcal{X}}$ induced by columns of Table 5.

whether or not xRy (we used this in Tables 2 and 5). Relation R may also be coded as a family of sets $\mathcal{X} = \{xR : x \in X\}$, where $xR = \{y \in Y : xRy\}$ for all $x \in X$, or, symmetrically, as a family of sets $\mathcal{Y} = \{Ry : y \in Y\}$, where $Ry = \{x \in X : xRy\}$ for all $y \in Y$.

Data of this kind are the basis of several significant paradigms in psychometrics and related areas. One is Guttman's "scalogram analysis of qualitative data" [Guttman, 1944, 1950]. In this paradigm, X may be a psychometric test made up of dichotomous questions, Y a sample of subjects, and R may specify which questions are correctly answered by which subjects. If family of sets $\mathcal{X}$ is linearly ordered by set inclusion (the "scalability hypothesis" in Guttman's theory), then a weak order $\preceq$ becomes determined over test X , by setting $x \preceq x'$ iff $xR \supseteq x'R$, for all $x, x' \in X$ (in the same condition, and through a dual rule, a weak order on the sample of subjects Y is also determined). Another paradigm is the theory of "biorders" [Ducamp, Falmagne, 1969; Doignon, Ducamp, Falmagne, 1984]. Terms X , Y and R have the same meanings as in Guttman's theory. The main theorem on biorders specifies conditions for the existence of real valued functions f on X and g on Y so that xRy iff $f(x) \leq g(y)$, for all $x \in X$ and $y \in Y$ – i.e., question x is correctly answered by subject y iff ability $g(y)$ of the subject is no less than difficulty $f(x)$ of the question. The existence of such functions implies that, on set $X \cup Y$, a weak order $\preceq$ is defined so that $x \preceq x'$ iff $xR \supseteq x'R$ for all $x, x' \in X$, and $y \preceq y'$ iff $Ry \subseteq Ry'$ for all $y, y' \in Y$, and these connections are the same as those in Guttman's model. One third paradigm deserving mention is the "Boolean analysis of questionnaires" [Flament, 1965, 1976]. In this case, set X is a questionnaire of dichotomous

$A_1 = \{y_1, y_2, y_5, y_7, y_8, y_{12}\}, B_1 = \{y_2, y_3, y_6, y_7, y_8, y_9, y_{11}, y_{13}, y_{14}\},$ $C_1 = \{y_2, y_3, y_8, y_9, y_{10}, y_{13}, y_{14}\}, D_1 = \{y_1, y_3, y_{12}\},$ $E_1 = \{y_1, y_2, y_3, y_5, y_6, y_7, y_8, y_{11}, y_{12}, y_{13}\}, F_1 = \{y_2, y_3, y_{11}, y_{13}, y_{14}\},$ $G_1 = \{y_2, y_3, y_7, y_8, y_9, y_{10}, y_{13}, y_{14}\}, H_1 = \{y_1, y_3, y_4, y_5, y_6, y_7, y_{11}, y_{12}, y_{13}, y_{15}\},$ $I_1 = \{y_1, y_2, y_5, y_6, y_8\}, J_1 = \{y_1, y_3, y_4, y_5, y_6, y_{13}, y_{15}\},$ $K_1 = \{y_1, y_3, y_{11}, y_{12}\}, L_1 = \{y_1, y_2, y_5, y_8, y_{12}\}$ $\delta(C; G), \delta(D; E, H, K), \delta(F; B), \delta(I; E), \delta(J; H), \delta(K; E, H), \delta(L; A, E)$ $A_8 = \{y_1, y_2, y_5, y_7, y_8, y_{12}\}, B_8 = \{y_2, y_3, y_6, y_7, y_8, y_9, y_{11}, y_{13}, y_{14}\}, C_8 = \emptyset, D_8 = \emptyset,$ $E_8 = \{y_1, y_2, y_3, y_5, y_6, y_7, y_8, y_{11}, y_{12}, y_{13}\}, F_8 = \emptyset, G_8 = \{y_2, y_3, y_7, y_8, y_9, y_{10}, y_{13}, y_{14}\},$ $H_8 = \{y_1, y_3, y_4, y_5, y_6, y_7, y_{11}, y_{12}, y_{13}, y_{15}\}, I_8 = \emptyset, J_8 = \emptyset, K_8 = \emptyset, L_8 = \emptyset$ $\rho(G, y_{10}), \rho(H, y_4), \rho(H, y_{15})$ $A_{11} = \{y_1, y_2, y_5, y_7, y_8, y_{12}\}, B_{11} = \{y_2, y_3, y_6, y_7, y_8, y_9, y_{11}, y_{13}, y_{14}\}, C_{11} = \emptyset, D_{11} = \emptyset,$ $E_{11} = \{y_1, y_2, y_3, y_5, y_6, y_7, y_8, y_{11}, y_{12}, y_{13}\}, F_{11} = \emptyset, G_{11} = \{y_2, y_3, y_7, y_8, y_9, y_{13}, y_{14}\},$ $H_{11} = \{y_1, y_3, y_5, y_6, y_7, y_{11}, y_{12}, y_{13}\}, I_{11} = \emptyset, J_{11} = \emptyset, K_{11} = \emptyset, L_{11} = \emptyset$ $\delta(A; E), \delta(G; B), \delta(H; E)$ $A_{14} = \emptyset, B_{14} = \{y_2, y_3, y_6, y_7, y_8, y_9, y_{11}, y_{13}, y_{14}\}, C_{14} = \emptyset, D_{14} = \emptyset,$ $E_{14} = \{y_1, y_2, y_3, y_5, y_6, y_7, y_8, y_{11}, y_{12}, y_{13}\},$ $F_{14} = \emptyset, G_{14} = \emptyset, H_{14} = \emptyset, I_{14} = \emptyset, J_{14} = \emptyset, K_{14} = \emptyset, L_{14} = \emptyset$ $\rho(B, y_9), \rho(B, y_{14})$ $A_{16} = \emptyset, B_{16} = \{y_2, y_3, y_6, y_7, y_8, y_{11}, y_{13}\}, C_{16} = \emptyset, D_{16} = \emptyset,$ $E_{16} = \{y_1, y_2, y_3, y_5, y_6, y_7, y_8, y_{11}, y_{12}, y_{13}\},$ $F_{16} = \emptyset, G_{16} = \emptyset, H_{16} = \emptyset, I_{16} = \emptyset, J_{16} = \emptyset, K_{16} = \emptyset, L_{16} = \emptyset$ $\delta(B; E)$ $A_{17} = \emptyset, B_{17} = \emptyset, C_{17} = \emptyset, D_{17} = \emptyset, E_{17} = \{y_1, y_2, y_3, y_5, y_6, y_7, y_8, y_{11}, y_{12}, y_{13}\},$ $F_{17} = \emptyset, G_{17} = \emptyset, H_{17} = \emptyset, I_{17} = \emptyset, J_{17} = \emptyset, K_{17} = \emptyset, L_{17} = \emptyset$ $\rho(E, y_1), \dots, \rho(E, y_{13})$ $A_{27} = \emptyset, B_{27} = \emptyset, C_{27} = \emptyset, D_{27} = \emptyset, E_{27} = \emptyset, F_{27} = \emptyset,$ $G_{27} = \emptyset, H_{27} = \emptyset, I_{27} = \emptyset, J_{27} = \emptyset, K_{27} = \emptyset, L_{27} = \emptyset$

Table 7. Evolution of the family of sets specified by Table 5 in a complete implementation of the GYO algorithm.

items, set Y is a sample of participants, and the Boolean matrix describing relation R is the complete record of answers in the experiment. Boolean analysis aims at discovering plausible “implications” between items in the questionnaire, i.e., logical connections like this: “if participants give positive answers to items x_i and x_j , then they also give positive answers to item x_h or item x_k (or both)”. This implication is accepted as well-supported by relation R obtained from the experiment, if the number of participants whose “response patterns” invalidate it is very small, i.e., if number $|x_i R \cap x_j R \cap (Y \setminus x_h R) \cap (Y \setminus x_k R)|$ is negligible. One fourth paradigm is “formal concept analysis” [Wille, 1982; Ganter, Wille, 1999]. In typical applications, X is a set of attributes, Y a set of objects, and xRy (for any $x \in X$ and $y \in Y$) means that object y is endowed with attribute x – system (X, Y, R) is called a “context”. One basic target of formal concept analysis is to identify “dependences”

FIGURE 3. Alternative partial tree representations of the qualitative proximity structure from Table 5.

between sets of attributes: if $U = \{u_1, \dots, u_h\}$ and $V = \{v_1, \dots, v_k\}$ are subsets of X , then dependence $U \rightarrow V$ is judged to be valid in the presumed context if $(u_1 R \cap \dots \cap u_h R) \subseteq (v_1 R \cap \dots \cap v_k R)$. Based on a network of dependences thus determined, an overall algebraic structure is obtained, involving both sets X and Y , and called the “concept lattice” inherent in the given context.

When compared with these and similar paradigms, the approach discussed in this paper distinguishes itself mainly in the type of information which is extracted from the data and becomes the direct subject of representation (another salient difference is our reference to trees as model structures). The thematic empirical information is a ternary relation on set X , i.e., the intersection-betweenness of Definition 1 (a relation of the same kind may be determined on set Y , by interchanging the roles of X and Y in the formula). We say that this ternary relation expresses qualitative (rather than quantitative) proximities because it is determined through comparisons involving sets (rather than numbers). Similar uses of the distinction “qualitative” vs “quantitative” are customary in the psychometric literature (e.g., [Restle, 1959; Tversky, 1977; Candel, 1997]).

There are models in psychometrics to which our proposal comes quite close, as regards the kind of thematic information. In particular, I mean the concepts of “linear array” as defined by Restle [1959, 1961] and “chain” as defined by Tversky, Gati [1982]. Both are framed within the “feature matching approach”, which presumes that a set X of “objects” and a set Y of “features” are given, and that each object $u \in X$ may be represented as a set of features $U \subseteq Y$, i.e., the set of features in Y that are possessed by object u . Set X is said to form a “linear array” if there is a total ordering $(u_1, \dots, u_n)$ of its elements so that for all $1 \leq i, j, k \leq n$:

$$i < j < k \text{ iff } U_i \cap U_k \subseteq U_j \subseteq U_i \cup U_k.$$

In a weaker – and more general – sense, set X is said to form a “chain” if there is a total ordering $(u_1, \dots, u_n)$ of its elements so that for all $1 \leq i, j, k \leq n$:

$$i < j < k \text{ iff } U_i \cap U_k \subseteq U_j.$$

Thus, the set-theoretic information which is involved in defining a “chain” is precisely intersection-betweenness (Definition 1). The information entering the definition of a “linear array” is stronger, as condition $U_i \cap U_k \subseteq U_j$ for a generic “chain” becomes integrated by condition $U_j \subseteq U_i \cup U_k$ (note that the combined condition

qualifying “linear arrays” abstractly resembles formula $x \wedge z \leq y \leq x \vee z$, which expresses one possible way of determining a betweenness relation within a lattice (cf. [Birkhoff, 1967, §II.9; Fishburn, 1971, §1]). Now, the most salient difference of our approach, compared with theorisations on the two above concepts, is that, in our representation problems, reference is made to tree graphs in general, rather than to linear orders or chain graphs in particular. Indeed, “chains”, as understood by Tversky and Gati, precisely correspond to those families of sets which allow for full chain representation – a special kind of full tree representation, considered at the end of Section 2 (Corollary 2). Another difference of our approach is that we discussed not only “full” but also “partial” tree representations of intersection betweennesses – the concept of “linear array” of Restle and that of “chain” of Tversky and Gati are both conditional on the existence of a linear order (or chain graph) which *fully* represents the ternary (betweenness) relation derived from the data. Such differences also distinguish our approach from certain uses of the betweenness concept in the theory of psychological measurement (e.g., see [Krantz, Luce, Suppes, Tversky, 1971, chap. 4], concerning betweenness relations in the analysis of “difference structures”).

Representation of proximities by means of tree graphs is a common characteristic of several methods of combinatorial descriptive psychometrics (cf. [Barthélemy, Guénoche, 1991]). Preference for trees as tools for efficient representation of data is quite understandable, considering that, among graph-theoretic structures, trees provide a reasonable balance between the opposite requirements of flexibility and simplicity: they are more flexible than mere chains, but may be viewed as the simplest and most easily readable graphs after chains. In order to locate our contribution within the variegated panorama of tree-using psychometric methods, the following circumstances may be considered. The information we considered for representation by trees is of a qualitative rather than quantitative type, in that it is derived from a family of sets (or a relation between sets) through Boolean comparisons, rather than from a table of proximity measures (a numerical index). In this respect, our proposal differs from standard methods using trees on numerical proximity structures (like hierarchical cluster analysis and additive similarity trees), but comes close to methods applying trees to non-numerical sorts of proximity information, like those discussed by Flament [1978], Defays [1979], Colonius, Schulze [1981], and Leclerc [1987].

A significant connection between our representation problems and the approach of Flament [1978] has already been discussed in Section 3 (Theorem 3). Here I wish to add a few comments on the relationship of my study with that of Defays [1979]. Besides referring to trees for representing qualitative proximities, both studies have in common this character: the proximity information to be represented is a ternary relation – a betweenness – on an empirical domain. However, there is one important difference: the betweenness relation considered by Defays is presumed to be determined through direct judgments by the subjects of an experiment (each judgment being the comparison of three objects in the empirical domain), whereas intersection-betweenness in this study derives from a family of sets (which, in turn, may derive from a relation between two basic sets). Thus, our betweenness has a peculiar set-theoretic descent, along a line which authorises comparisons with the several models mentioned above in this section (Guttman, Restle, Tversky, etc.).

Similar comparisons are not plausible, or at least not obvious for the Defays' constructs. There are another similarity and another difference worthy of mention. In both Defays' and this study, conditions are sought first for "full" and then for "partial" tree representation of a betweenness relation. However, "partiality" is differently understood in the two approaches: in Defays' study partiality concerns the relationship between objects in the empirical domain and vertices in the representing tree, whereas in this study partiality concerns the relationship between triples in the empirical betweenness and triples in the betweenness of the representing tree.

REFERENCES

- ABITEBOUL S., HULL R., VIANU V., *Foundations of databases*, Reading (MA), Addison-Wesley, 1995.
- BARTHÉLEMY J.-P., GUÉNOCHE A., *Trees and proximity representations*, Chichester, Wiley, 1991.
- BEERI C., FAGIN R. MAIER D., YANNAKAKIS M., "On the desirability of acyclic database schemes", *Journal of the Association of Computing Machinery* 30, 1983, p. 479-513.
- BERNSTEIN P.A., GOODMAN N., "Power of natural semijoins", *SIAM Journal on Computing* 10, 1981, p. 751-771.
- BIRKHOFF G., *Lattice theory*, Providence (RI), American Mathematical Society, 1967.
- CANDEL M.J.J.M., "Exploring the set-theoretical structure of objects by additive trees", *Psychometrika* 62, 1997, p. 119-131.
- COLONIUS H., SCHULZE H.H., "Tree structures for proximity data", *British Journal of Mathematical and Statistical Psychology* 34, 1981, p. 167-180.
- DEFAYS D., "Tree representations of ternary relations", *Journal of Mathematical Psychology* 19, 1979, p. 208-218.
- DOIGNON J.-P., DUCAMP A., FALMAGNE J.-C., "On realizable biorders and the biorder dimension of a relation", *Journal of Mathematical Psychology* 28, 1984, p. 73-109.
- DUCAMP A., FALMAGNE J.-C., "Composite measurement", *Journal of Mathematical Psychology* 6, 1969, p. 359-390.
- FISHBURN P.C., "Betweenness, orders and interval graphs", *Journal of Pure and Applied Algebra* 1, 1971, p. 159-178.
- FLAMENT C., *L'analyse booléenne de questionnaire*, Paris, Mouton, 1976.
- FLAMENT C., "Hypergraphes arborés", *Discrete Mathematics* 21, 1978, p. 223-227.
- FLAMENT C., "L'analyse booléenne de questionnaire", *Mathématiques et Sciences humaines* 12, 1965, p. 3-10.
- GANTER B., WILLE R., *Formal concept analysis: Mathematical foundations*, Berlin, Springer-Verlag, 1999.
- GATI I., TVERSKY A., "Representations of qualitative and quantitative dimensions", *Journal of Experimental Psychology: Human Perception and Performance* 8, 1982, p. 325-340.
- GRAHAM M.H., *On the universal relation*, Technical Report, University of Toronto, Toronto (Ontario), 1979.
- GUTTMAN L., "A basis for scaling qualitative data", *American Sociological Review* 9, 1944, p. 139-150.

- GUTTMAN L., "The basis for scalogram analysis", S.A. Stouffer, L. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star, J.A. Clausen (eds.), *Measurement and prediction*, New York, Wiley, 1950, p. 60-91.
- KRANTZ D.H., LUCE R.D., SUPPES P., TVERSKY A., *Foundations of measurement*. Volume 1. *Additive and polynomial representations*, New York, Academic Press, 1971.
- LECLERC B., "Arbres minimums communs et compatibilité de données de types variés", *Mathématiques et Sciences humaines* 98, 1987, p. 41-67.
- MAIER D., *The theory of relational databases*, Rockville (MD), Computer Science Press, 1983.
- PITCHER E., SMILEY M.F., "Transitivities of betweenness", *Transactions of the American Mathematical Society* 52, 1942, p. 95-114.
- RESTLE F., *Psychology of judgment and choice: A theoretical essay*, New York, Wiley, 1961.
- RESTLE F., "A metric and an ordering on sets", *Psychometrika* 24, 1959, p. 207-220.
- SHOLANDER M., "Trees, lattices, order, and betweenness", *Proceedings of the American Mathematical Society* 3, 1952, p. 369-381.
- TVERSKY A., "Features of similarity", *Psychological Review* 84, 1977, p. 327-352.
- TVERSKY A., GATI I., "Similarity, separability, and the triangle inequality", *Psychological Review* 89, 1982, p. 123-154.
- WILLE R., "Restructuring lattice theory: An approach based on hierarchies of concepts", I. Rival (ed.), *Ordered sets*, Dordrecht, Reidel, 1982, p. 445-470.
- YU C.T., OZSOYOGLU M.Z., "An algorithm for tree-query membership of a distributed query", *Proceedings 1979 IEEE COMPSAC*, New York, Institute of Electrical and Electronic Engineers, 1979, p. 306-312.