A pretopological approach for structural analysis

Christine Largeron *, Stéphane Bonnevay

Université Claude Bernard Lyon I, UMR CNRS 5823-bât 101-43, Boulevard du 11/11/1918, 69 622 Villeurbanne Cedex, France

Abstract

The aim of this paper is to present a methodological approach for problems encountered in structural analysis. This approach is based upon the pretopological concepts of pseudoclosure and minimal closed subsets. The advantage of this approach is that it provides a framework which is general enough to model and formulate different types of connections that exist between the elements of a population. In addition, it has enabled us to develop a new structural analysis algorithm. An explanation of the definitions and properties of the pretopological concepts applied in this work is first shown and illustrated in sample settings. The structural analysis algorithm is then described and the results obtained in an economic study of the impact of geographic proximity on scientific collaborations are presented. © 2002 Elsevier Science Inc. All rights reserved.

1. Introduction

The data of a structural analysis problem are represented by a finite set E, composed of elements which are related to each other by some form of connection. The goal of structural analysis is to highlight groups of "interdependent" elements. The necessity of applying precise concepts to illustrate notions

---

* Corresponding author. Tel.: +33-4-72-43-16-57; fax: +33-4-72-44-05-73.

  E-mail addresses: largeron@univ-st-etienne.fr (C. Largeron), bonnevay@univ-lyon1.fr (S. Bonnevay).

as abstract as "interdependency" and "connection" has been recognized for years in structural analysis [11,17]. Indeed, it is possible to perceive connections between elements by means of dissimilarity measures, binary relations, or neighborhoods. It is also possible to define a graph structure, valued or not, on the population. Depending on the concept retained to formulate the connection between the elements of a population, structural analysis problems can be approached in various manners: from a metric point of view if distance has been retained, in terms of topological space if neighborhoods have been chosen, or by graph theory.

In this paper, we will apply the concepts of pseudoclosure and minimal closed subsets that have been developed in pretopological theory [2,9]. The advantage of this approach is that it enables us to formulate and treat structural analysis problems in a unified manner, even if the connections between elements are diverse. It is enough to select a pseudoclosure adapted to the application. Obviously, according to certain definitions of pseudoclosures, our approach provides results which can also be obtained using usual methods like graph algorithms or single linkage method.

Closure operators have been widely studied in algebra [3], topology [12] and computer science theory. Nevertheless, the axiomatics that define them are often too limited to treat concrete problems. For this reason, in pretopology [2], as in [7,8], it is not assumed that the pseudoclosure application is idempotent. This approach enabled us to develop a new algorithm for structural analysis [4,5,13]. The principle consists in constructing a pseudoclosure application $a(\cdot)$ from parts of E onto themselves based on the connection between the elements of population E. This pseudoclosure expresses the extension phenomena (dilatation, propagation, influence, etc.) of these subsets. As the pseudoclosure application is not idempotent, it is the successive aggregations that lead to the obtention of closed subsets. These closed subsets represent the homogenous or interdependent subsets in relation to the pseudoclosure function. The algorithm determines the smallest possible closed subsets of E (minimal closed subsets), then those which contain them, until the structural analysis of the entire population has been completed. This means defining an inclusion relation on the set of minimal closed subsets.

It is important to note that the structural method is not a clustering method: it does not build a partition or a hierarchy on the population E. In the structuration process, the aim is:

- first to define the closed subset associated to each element x of E: this subset contains all elements of E related directly or not directly to x,
- secondly to display the inclusion relation between these closed subsets.

Thus, the aim of our method is to uncover the structure of the population E (relations between elements, relations between groups of elements, etc.) with a view to extracting new information from E.

The pretopological concepts applied in this study are defined in the following section. They are illustrated by a simple geographical example on artificial data. The structural analysis algorithm is then presented. In order to illustrate the method, an application to co-authorships of publications between French geographical areas is provided.

2. The pretopological concepts of "pseudoclosure" and "closure"

Let us consider a population E, composed of elements in the widest sense of the term (individuals, plots of land, etc.). E is a non-empty finite set, and $\mathbb{P}(E)$ designates all of the subsets of E.

2.1. Pseudoclosure and pretopological space definitions

Definition 1. A pseudoclosure is a map $a(\cdot)$ from $\mathbb{P}(E)$ to $\mathbb{P}(E)$ satisfying the following two conditions:

$(\text{P1}) : a(\emptyset) = \emptyset$

$(\text{P2}) : \forall A \in \mathbb{P}(E), \quad A \subset a(A)$

A pretopological space is a pair (E, a) where E is endowed with a pseudoclosure $a(\cdot)$ [2].

The subset $a(A)$ is also called pseudoclosure of A. As said before, $a(a(A))$ is not always equal to $a(A)$. Thus, $a(\cdot)$ can be applied on a set A in sequence, so as to model expansions: $A \subset a(A) \subset a^2(A) \subset \cdots$.

This definition of pretopological space is not exactly the same in other works like in [7] where the application $a(\cdot)$ must verify not only (P1) and (P2) but the following as well:

$(\text{P3}) : \forall A \in \mathbb{P}(E), \ \forall B \in \mathbb{P}(E), \quad a(A \cup B) = a(A) \cup a(B).$

In [2], this is the definition of a particular space which is called D-pretopological space.

Definition 2. A V-pretopological space (E, a) is a pretopological space that satisfies (P4):

$(\text{P4}) : \forall A \in \mathbb{P}(E), \ \forall B \in \mathbb{P}(E), \quad A \subset B \Rightarrow a(A) \subset a(B)$

A D-pretopological space is necessarily a V-pretopological space but the converse property is false $((\text{P3}) \Rightarrow (\text{P4}))$. The demonstration of this can be seen in [2]. Throughout the remainder of this paper (E, a) is at least a V-pretopological space which is less restrictive than a D-pretopological space. In this

context of V-space, the pretopological approach is really useful especially to formulate connection between elements in real applications when usual methods are not suitable.

2.2. Construction of a V-pretopological structure based on a pseudoclosure

There are of course many ways in which a pseudoclosure can be constructed from, either real data, or from other properties E is equipped with. In order to illustrate the concepts, we will give three short examples where E is composed of 10 elements. We suppose that each element of E represents a plot of land in a given geographical zone, each of which contains a form of water supply (aquifer, spring, etc.). We will study the risk of toxic products spreading across the plots of land. The connections between the elements of E will be modeled by a distance (case 1), a binary relation (case 2), and a valued graph (case 3) depending on the manner in which these products are disseminated

Case 1, where E is endowed with a metric defined by a distance d.
Let r be a positive real. For each element x of E, $B(x,r)$ is defined by: $B(x,r) = \{y \in E, d(x,y) \leqslant r\}$.
A pseudoclosure $a(\cdot)$ can be defined from $B(x,r)$ by:

$$\forall A \in P(E), \quad a(A) = \{x \in E, B(x,r) \cap A \neq \emptyset\} \tag{A1}$$

$a(A)$ is composed of all elements of A and all elements $y \notin A$ such as y is not so far (with regard to r) from, at least, one element of A.

It is clear that application $a(\cdot)$, thus defined, verifies axioms (P1), (P2) and also (P3) and (P4), and as such, the pretopological structure induced by $a(\cdot)$ on E is Type D and thus also Type V.

Example 1. Table 1 indicates, for each element x of E, the coordinates $(x1, x2)$ of x in the plane, $B(x, 2)$, the pseudoclosure $a(x)$ of x. This case could correspond to the situation where a plot y can be contaminated by plot x, if x and y are geographically situated close by to one another, i.e., if they are situated at a distance that is less or equal than a specific predetermined level r (e.g. $r = 2$).

Case 2, where E is endowed with a binary and reflexive relation.
Let R be a reflexive binary relation defined on E. R is not necessarily symmetric. We define $R(x) = \{y \in E, xRy\}$ and $R^{-1}(x) = \{y \in E, yRx\}$.

As the relation R is reflexive, x belongs to $R(x)$ and to $R^{-1}(x)$.

From a such relation R, it is possible to define several D-pretopological spaces with pseudoclosure satisfying for example (A2) or (A3):

$$\forall A \in P(E), \quad a(A) = \{x \in E, R(x) \cap A \neq \emptyset\}, \tag{A2}$$

$$\forall A \in P(E), \quad a(A) = \{x \in E, R^{-1}(x) \cap A \neq \emptyset\}. \tag{A3}$$

Both these pseudoclosures are not equivalent when R is not symmetric.

Table 1

x	$x1$	$x2$	$B(x,2) = \{y \in E, d(x,y) \leqslant 2\}$	$a(x) = \{y \in E, B(y,2) \cap \{x\} \neq \emptyset\}$
1	1	2	$\{1,2\}$	$\{1,2\}$
2	3	2	$\{1,2,3,4,5\}$	$\{1,2,3,4,5\}$
3	3	4	$\{2,3,6\}$	$\{2,3,6\}$
4	4	2	$\{2,4,5\}$	$\{2,4,5\}$
5	4	1	$\{2,4,5\}$	$\{2,4,5\}$
6	5	4	$\{3,6,7\}$	$\{3,6,7\}$
7	6	4	$\{6,7\}$	$\{6,7\}$
8	8	3	$\{8,9,10\}$	$\{8,9,10\}$
9	9	2	$\{8,9,10\}$	$\{8,9,10\}$
10	8	1	$\{8,9,10\}$	$\{8,9,10\}$

$a(A)$ is composed of A (R is reflexive) and all elements y in relation (R or R^{-1}) with, at least, one element of A.

Example 2. Table 2 illustrates, for all elements x of E, set $R(x)$ and its pseudoclosure $a(x)$ verifying (A2). In this example, the connection between the elements of E could be formulated by a binary reflexive relation R, which is not symmetric, such that for all pairs (x,y) belonging to $E \times E$, we have xRy, if there is a pipe which conveys products from y to x.

Case 3, *where* (E,U,L) *is a valued and directed graph.*
  Let (E,U,L) be a directed and valued graph, with:
  E: the set of vertices;
  U: the set of edges ($U \subset \{(x,y)/x \in E, y \in E\}$);
  L: a valuation map which associate to $u \in U$ the length $L(u)$.
  The pseudoclosure $a(\cdot)$ can be defined by

$$\forall A \in \mathbb{P}(E), \quad a(A) = \left\{ y \in E - A \middle/ \sum_{x \in A} L(x,y) \geqslant s \right\} \cup A, \tag{A4}$$

where s is a positive real.

Table 2

x	$R(x) = \{y \in E/xRy\}$	$a(x) = \{y \in E, R(y) \cap \{x\} \neq \emptyset\}$
1	$\{1\}$	$\{1,2,3\}$
2	$\{1,2\}$	$\{2,3,4,5\}$
3	$\{1,2,3,7\}$	$\{3,6\}$
4	$\{2,4\}$	$\{4\}$
5	$\{2,5\}$	$\{5\}$
6	$\{3,6\}$	$\{6,7\}$
7	$\{6,7,8\}$	$\{3,7\}$
8	$\{8,9\}$	$\{7,8,9,10\}$
9	$\{8,9\}$	$\{8,9\}$
10	$\{8,10\}$	$\{10\}$

$a(A)$ is composed of A and all elements y where the sum of valued edges between some elements of A and y is greater than the threshold s.

It is easy to verify that (E, a) is a V-pretopological space but it is not a D-pretopological space ((P3) is not satisfied).

Example 3. Table 3 gives the value $L(u)$ of each arc u of U, whereas Table 4 provides the pseudoclosure $a(x)$. In this example, the connection between the elements of E, taken two by two, can be represented by a valued graph $G = (E, U, L)$ where for each pair of vertices $(x, y) \in E \times E$, there is an edge of x towards y, if there is a pipe which conveys products from x to y. The valuation $L(x, y)$ corresponds to the capacity of the flow between plots x and y. We will consider here that a plot y can only be contaminated by plots for which there is a pipe which conveys products from y to x and for which the sum of the flows arriving at x is equal to or greater than a given level s (e.g. $s = 2$).

Table 3

$u \in U$	$L(u)$
$(1, 2)$	2
$(1, 3)$	1
$(2, 3)$	1
$(2, 4)$	1
$(2, 5)$	2
$(3, 6)$	2
$(6, 7)$	1
$(7, 3)$	2
$(8, 7)$	2
$(8, 9)$	3
$(8, 10)$	1
$(9, 8)$	3

Table 4

x	$a(x)$
1	$\{1, 2\}$
2	$\{2, 5\}$
3	$\{3, 6\}$
4	$\{4\}$
5	$\{5\}$
6	$\{6\}$
7	$\{3, 7\}$
8	$\{7, 8, 9\}$
9	$\{8, 9\}$
10	$\{10\}$

2.3. Closed subsets, minimal closed subsets and elementary closed subsets

Contrary to topological theory, in pretopology, a pseudoclosure is not idempotent: for a given subset A of E, $a(A)$ may be contained in $a(a(A))$ without being equal to $a(a(A))$:

$$\forall A \subset E, \quad a(A) \subset a(a(A)).$$

In fact, the property of idempotence is verified only by the so-called closed subsets of E. These closed subsets are of particular interest within the context of structural analysis. They enable the representation of the homogenous subsets of E in regard to the pseudoclosure retained. A closed subset F is a subset of E and for which no elements of $E–F$ belong to the pseudoclosure of F.

More precisely, given a set E endowed with a V-pretopological structure defined by a pseudoclosure $a(\cdot)$, let us recall the following definitions pertaining to closed subset.

Definition 3. A subset F of E such as $a(F) = F$ is called a closed subset of E for $a(\cdot)$.

Let $\Im(E, a)$ be the family of closed subsets of E for $a(\cdot)$:

$$\Im(E, a) = \{F \subset E / a(F) = F\},$$

and

$$\Im(E, a)^* = \Im(E, a) - \{\emptyset\}.$$

Property 1. *In a V pretopological space the intersection of closed subsets is a closed subset.*

Proof. Given, $A = \bigcap_{i \in I} A_i$, with A_i closed subset.
  We first prove that $a(A) \subset A$:
  By definition of $A : \forall i \in I, \; A \subset A_i$:
  In a Type V pretopological space: $\forall i \in I, \; a(A) \subset a(A_i)$,
  As A_i is a closed subset: $\forall i \in I, \; a(A_i) = A_i$,
  So: $\forall i \in I, \; a(A) \subset A_i$,
  And $a(A) \subset \bigcap_{i \in I} A_i \iff a(A) \subset A$,
  By definition of a pseudoclosure: $A \subset a(A)$,
  Thus $A = a(A)$ Therefore A is a closed subset. $\square$

Definition 4. Let F be a subset of E. The closure of F is the smallest closed subset in terms of inclusion in the family $\Im(E, a)$, containing F.

Property 2. *In a Type V pretopological space, each subset of E possesses a closure.*

Proof.

Given $A \in \mathrm{P}(E)$ and $\mathfrak{I}_A$ the set of closed sets containing $A : \mathfrak{I}_A = \{G \in \mathfrak{I}(E, a) / A \subset G\}$.

As E is a closed set, $\mathfrak{I}_A \neq \emptyset$ because $E \in \mathfrak{I}_A$.

Given $H_A = \bigcap_{G \in \mathfrak{I}_A} G$. H_A is a closed subset according to the property (P1).

Thus, the smallest closed subset containing A, noted F_A, is contained in H_A, but as H_A is the intersection of all closed subsets containing A, then: $F_A = H_A$. The closure of a subset A included in E is therefore equal to the intersection of all closed subsets containing A. □

Property 2 being true for all subsets of E, it is particularly true for all singletons.

Definition 5. An elementary closed subset, noted as F_x, is the closure of a one element set $\{x\}$ of E.

Note that $\mathfrak{I}e(E, a)$ represents the set of elementary closed subsets of E: $\mathfrak{I}e(E, a) = \{F_x, x \in E\}$, therefore:

$$\mathfrak{I}e(E, a) \subset \mathfrak{I}(E, a)^*.$$

Property 3. *Two distinct elementary closed subsets F_x and F_y are either disjoint $(F_x \cap F_y = \emptyset)$ or contain a non-empty intersection such that for all $z \in F_x \cap F_y$, we have $F_z \subset F_x \cap F_y$ (see Fig. 1).*

Proof. Let $z \in F_x \cap F_y$ and let F_z be the smallest closed subset containing z in the sense of inclusion in $\mathfrak{I}(E, a)$. According to Property 1, $F_z \cap F_x$ is a closed subset, therefore: $F_z \subset F_x \cap F_y$. □

Definition 6. A minimal closed subset of E with regard to $a(\cdot)$ is an element of $\mathfrak{I}(E, a)^*$ minimal in terms of inclusion in $\mathfrak{I}(E, a)^*$.

Let $\mathfrak{I}m(E, a)$ represent the set of minimal closed subsets of E:

$$\mathfrak{I}m(E, a) = \{F \in \mathfrak{I}(E, a)^*, \neg(G \in \mathfrak{I}(E, a)^* - \{F\}, G \subset F)\}.$$
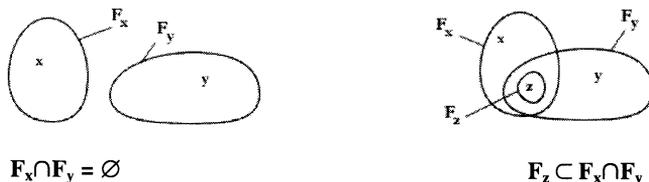

Fig. 1. Property 3.

Table 5

x	Case 1 F_x	Case 2 F_x	Case 3 F_x
1	$\{1, 2, 3, 4, 5, 6, 7\}^*$	$\{1, 2, 3, 4, 5, 6, 7\}$	$\{1, 2, 3, 5, 6\}$
2	$\{1, 2, 3, 4, 5, 6, 7\}^*$	$\{2, 3, 4, 5, 6, 7\}$	$\{2, 5\}$
3	$\{1, 2, 3, 4, 5, 6, 7\}^*$	$\{3, 6, 7\}^*$	$\{3, 6\}$
4	$\{1, 2, 3, 4, 5, 6, 7\}^*$	$\{4\}^*$	$\{4\}^*$
5	$\{1, 2, 3, 4, 5, 6, 7\}^*$	$\{5\}^*$	$\{5\}^*$
6	$\{1, 2, 3, 4, 5, 6, 7\}^*$	$\{3, 6, 7\}^*$	$\{6\}^*$
7	$\{1, 2, 3, 4, 5, 6, 7\}^*$	$\{3, 6, 7\}^*$	$\{3, 6, 7\}$
8	$\{8, 9, 10\}^*$	$\{3, 6, 7, 8, 9, 10\}$	$\{3, 6, 7, 8, 9\}$
9	$\{8, 9, 10\}^*$	$\{3, 6, 7, 8, 9, 10\}$	$\{3, 6, 7, 8, 9\}$
10	$\{8, 9, 10\}^*$	$\{10\}^*$	$\{10\}^*$

According to [2], if E is finite, the existence of minimal closed subsets is guaranteed.

Examples: Table 5 indicates for each element x of E the elementary closed subset F_x, in the three previous cases. Minimal elementary closed subsets are designated by an asterisk $(*)$.

The following property proves that minimal closed subsets of $\Im(E, a)^*$ can be discovered in the elementary closed subsets $\Im e(E, a)$.

Property 4. $F \in \Im m(E, a) \Longleftrightarrow F \in \Im e(E, a)$ *and* F *is minimal by inclusion in* $\Im e(E, a)$.

Proof. In fact, it is enough to prove that:

(i) $F \in \Im m(E, a) \Rightarrow F \in \Im e(E, a)$ and F is minimal by inclusion in $\Im e(E, a)$:
Given $F \in \Im m(E, a)$, a minimal closed subset for inclusion in $\Im(E, a)^*$
Given $x \in F$, then $F_x \subset F$, for F_x is the smallest closed subset containing x. However, as F is minimal: $F_x = F$.
Therefore: $\forall F \in \Im m(E, a)$, $F \in \Im e(E, a)$ and F minimal for inclusion in $\Im(E, a)^*$.

(ii) $F \in \Im e(E, a)$ and F is minimal by inclusion in $\Im e(E, a) \Rightarrow F \in \Im m(E, a)$
If $F \notin \Im m(E, a)$ then $G \in \Im(E, a)^* - \{F\}$ exists such that $G \subset F$. However, $\exists y \in G$ therefore, $F_y \subset G \subset F$, which contradicts the fact that F is minimal in $\Im e(E, a)$. Therefore: $F \in \Im m(E, a)$. □

As the aim of the structural process is to find minimal closed subsets, it is easy to understand that this property reduces the algorithmic complexity.

3. Structural analysis process

The underlying idea of the structural analysis method is to first highlight homogenous groups (minimal closed subsets), then those containing them

(non-minimal elementary closed subsets) until the structural analysis of the entire population has been completed. Nevertheless, according to Property 4, it is enough to define the inclusion relation on the set of elementary closed subsets: $(\Im e(E,a), \subset)$. It is necessary to proceed in three stages:

- The first step consists in determining the set of elementary closed subsets $\Im e(E,a)$ by associating a closure F_x to all elements x of E by means of the function *ElementaryClosedSubsets* described hereafter.
- The second step aims at searching for minimal closed subsets $\Im m(E,a)$ by means of the function *MinimalClosedSubsets*. In line with the previous statement, this means enumerating the set of elementary minimal closed subsets by inclusion in $\Im e(E,a)$ (see Property 4).
- The third step is the structural analysis phase. The aim of this step is to picture the inclusion relation between elements of $\Im e(E,a)$. This process enables us to generate the structure from each elementary closed subset by means of successive enlargements.

The structural analysis which is named *StructuralAnalysis* can thus be defined as follows.

The inputs of our *StructuralAnalysis* procedure are:
- the population E,
- the pseudoclosure $a(\cdot)$ defined on E.

The outputs are:
- the family of the elementary closed subsets $\Im e(E,a)$,
- the family of the minimal closed subsets $\Im m(E,a)$,
- the structure characterized by relations of inclusion between minimal closed subsets and elementary ones and relations of inclusion between elementary closed subsets with each other.

```
Procedure StructuralAnalysis;
  begin
  // Computation of Ie(E, a) by associating a closure Fx to all elements x of
  E
  Ie(E, a) = ElementaryClosedSubsets(E);
  // Computation of Im(E, a) finding in Ie(E, a)
  Im(E, a) = MinimalClosedSubsets(Ie(E, a));
  // Extraction of the structure Ie(E, a)
  ExtractStructure(Ie(E, a), Im(E, a));
  end;
Function ElementaryClosedSubsets(E: set): set;
  var    F: subset;
         x: element;
  begin
  I = ∅;
  for all x ∈ E do
```

```
    begin
    F = a({x});
    while (a(F) ≠ F) do F = a(F);
    ℑ = ℑ ∪ {F};
    end;
  ElementaryClosedSubsets = ℑ;
end;
  Function MinimalClosedSubsets (ℑ: set): set;
    var    minimal: boolean;
            ℑm, ℑt: set;
            F, G: subset;
    begin
    ℑm = ∅;
    while (ℑ ≠ ∅) do
     begin
     let F ∈ ℑ;
     ℑ = ℑ − {F};
     minimal = true;
     ℑt = ℑ;
     while ((ℑt ≠ ∅) and (minimal)) do
       begin
       let G ∈ ℑt;
       if (G ⊂ F) then minimal = false // F isn't minimal
              else if (F ⊂ G) then ℑ = ℑ − {G}; // G isn't minimal
       ℑt = ℑt − {G};
       end;
     if(minimal) then ℑm = ℑm ∪ {F};
     end;
    MinimalClosedSubsets = ℑm;
    end;
  Procedure ExtractStructure (ℑe, ℑm: set);
    var   Q: queue;
           ℑ: set;
           F, G: subset;
    Begin
    Q = ∅;
    for all F ∈ ℑm do enqueue(Q,F);
    while (Q ≠ ∅) do
     begin
     F = dequeue(Q);
     ℑ = {G ∈ ℑe, F ⊂ G and F ≠ G}; // supersets to F
     for all G∈MinimalClosedSubsets (ℑ) do
       begin
```

       if G∉Q **then** enqueue(Q,G);
       G is a descendant of F;
       end;
     end;
     end;

Software can be found at http://lass.univ-lyon1.fr/softs/index.html

    Examples: The results of the structural analysis obtained by the algorithm are respectively illustrated in Figs. 2 (Case No. 1), 3 (Case No. 2) and 4 (Case No. 3).

    In the first case, the structural analysis brings two elementary closed subsets to light. In terms of toxic spread, we can deduce that if a plot of land belonging to an elementary closed subset is contaminated, then all of the other plots of land belonging to this elementary closed subset will also be contaminated. Furthermore, as the intersection of these closed subsets taken two by two is empty, there is no risk of toxic spread between elements belonging to distinct elementary closed subsets.
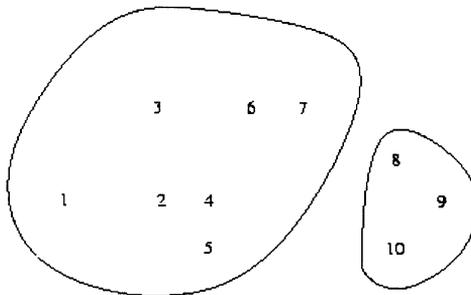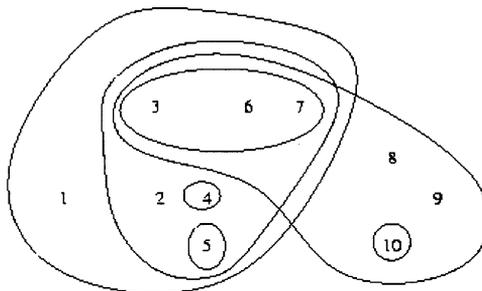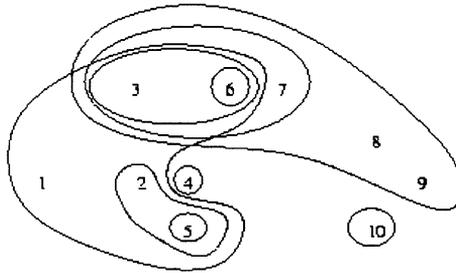

Fig. 2. Case 1.

Fig. 3. Case 2.

Fig. 4. Case 3.

The two other structuration cases can be interpreted as follows:

- an elementary closed subset F_x reduced to a single-element set $\{x\}$ corresponds to a plot of land x that cannot contaminate any other land in the geographical zone. This refers to elements $\{4\}$, $\{5\}$ and $\{10\}$ in Case No. 2, and $\{4\}$, $\{5\}$, $\{6\}$ and $\{10\}$ in Case No. 3;
- two elementary closed subsets F_x and F_y distinct such that $F_x \subset F_y$ correspond to two plots x and y, whereby toxic products may spread from y to x. In other words, if y is contaminated, then x will also be: either directly, or indirectly through a third plot of land;
- two elementary closed subsets F_x and F_y such that $F_x = F_y$ correspond to two plots x and y, whereby toxic products may spread from y to x or from x to y. Moreover, x and y can be contaminated by the same set of plots and can also contaminate the same other plots.

In the second case, for example, if any one plot of land $\{3, 6, 7\}$ is contaminated, the other two will also be, for $F_3 = F_6 = F_7$, and if 9 is contaminated, then 3 will also be, for $F_3 \subset F_9$.

As mentioned in Section 1, our method provides well known results in certain cases. For instance, when the pseudoclosure is defined by (A1) (case one), our process gives the partition obtained by the single-linkage method cutting the hierarchy when a marginal gain is greater or equal to the radius r of $B(x, r)$. In the same way, if E contains a pretopological structure defined in accordance with a pseudoclosure application verifying (A2), then the elementary closed subset associated with an element x of E is the set of successors of x in graph $G = (E, \sigma)$ where σ, the application which gives for each vertex x of E its successors, is such that $\sigma(x) = R(x)$. Thus, the set of elementary closed subsets $\Im e(E, a)$ in a pretopological space (E, a) where $a(\cdot)$ verifies (A2), corresponds to the transitive closure of σ in graph $G = (E, \sigma)$ where $\sigma(x) = R(x)$ for all elements x of E. It is obvious that in certain cases, general structural analysis method corresponds to less effective procedures (from an algorithmic point of view) than more specific algorithms. On the other hand, it provides a general framework which is applicable to a wider range of structural analysis

problems, especially when the pseudoclosure used to formulate connections between elements defines a V-pretopological space.

4. Application

To illustrate our method, we present the results of an economics study that analyses the impact of geographic proximity on scientific collaborations in the French context [1,6,14]. The question is: are scientific interactions favoured by geographic proximity? In order to answer this question, we use data on co-authored scientific publications between French "counties" (counties are French administrative geographical areas). Indeed, co-authorship is a good indicator of scientific interactions in scientometrics analysis. The data come from OST (Observatoire des Sciences et des Techniques) and are extracted from the Science Citation Index (SCI). For each year, we have a matrix $C = [c_{xy}]_{x,y \in \{1,\ldots,n\}}$, where c_{xy} gives the number of co-authored publications written by at least one author belonging to department x and at least one author belonging to county y and n is the total number of counties. The structural method enables us to show relations between counties according to co-authorhips of scientific publications. In that case, the connections between the counties are defined by a reflexive binary relation R such that $\mathrm{R}(x) = \{y \in E, x\mathrm{R}y\}$ is the set of counties with which x mainly publish:

$$\mathrm{R}(x) = \{y \in E, c_{xy} = \max\{c_{xz}, \mathrm{z} \neq \mathrm{x}\}\} \cup \{\mathrm{x}\}$$

R is not symmetric. The pseudoclosure is defined by (A2) ($\forall A \in \mathrm{P}(E)$, $a(A) = \{x \in E, \mathrm{R}(x) \cap A \neq \emptyset\}$) as in Section 2. So, by definition of R, a county x belongs to the pseudoclosure $a(A)$ of a set of counties A, if and only if x has mainly published with one county in A. $F(x)$ is the set of counties which have mainly published either with x directly or with other elements which have mainly published with x, directly or indirectly.

The result obtained for the matrix of co-authored publications in 1997 is illustrated in Fig. 5. An elementary closed subset $F(x)$ reduced to a singleton $\{x\}$ corresponds to a county such that $F(x) = a(\{x\}) = \{x\}$. It means that it does not exist any county that has mainly published with x. This refers for example to elements 61 and 62 (Orne and Pas-de-Calais). An element y, such that $y \in F(x)$, with $x \in E - \{y\}$, corresponds to a county y which has connections with x, either directly, or indirectly through other counties. For example, county 52 (Haute-Marne) has mainly published with 60 (Oise), so $52 \in F(60)$. In the same way, 30, 66, 84, 82 $\in F(34)$ and 82, 16, 24, 40, 64 and $101 \in F(33)$.

One notices that Paris (75) and Essonne (91) are very strong attractors as they include the whole set of other counties. It means all counties have directly or indirectly published with Paris and Essonne. Globally the result is consti-
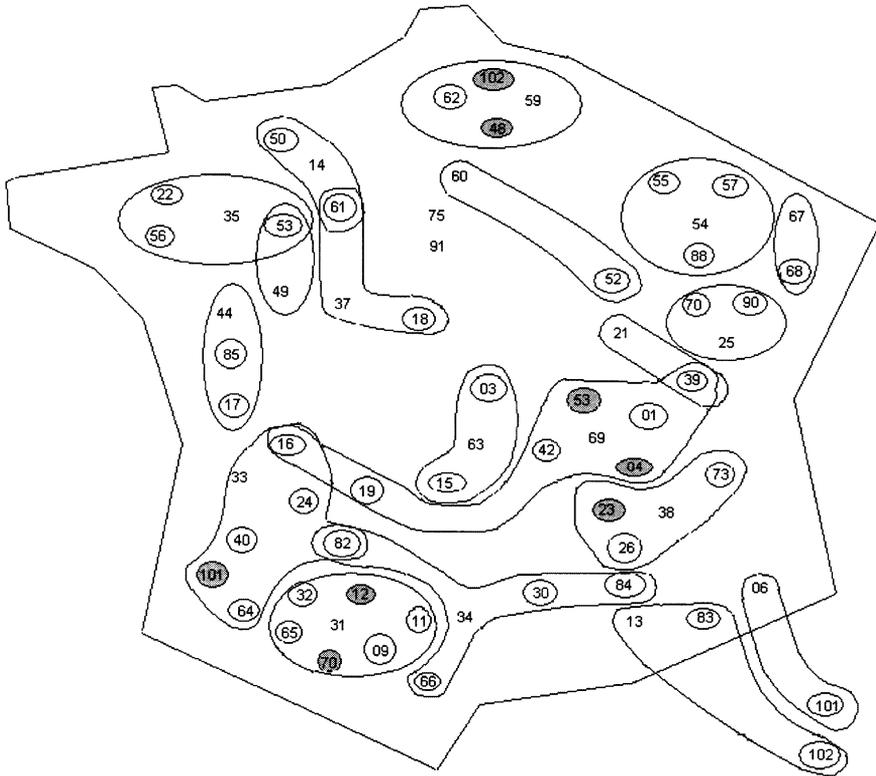
Fig. 5. Structuration of the co-authorship between French counties in 1997 ("departements" are situated in their correct geographic positions apart from those shaded in grey. They are identified by their correct post code number, except Haute corse (101) and Corse du Sud (102)).

tuted of separated groups around an attractor (single element inside the group), i.e. an element with which the other members of the group mainly publish. These attractors correspond to the main French University centres. They generally attract the smallest counties that are located close to them. Indeed, the counties with large universities do not privilege the relations between them: they publish inside the county or choose Paris as partner. They are, on the other hand, selected as main partners by the counties that surround them. Finally, counties having rather few publishing activities, carry these out with outside help primarily with the closest large universities. Thus, plotted on a map, this result shows clearly groups (sometimes connected to each other) around the main universities (Lille, Strasbourg, Clermond-Ferrand, Toulouse, Bordeaux, Lyon, Grenoble, Marseille, Nice and Montpellier). In the French case, globally, scientific interaction is favoured by geographic proximity even if

the role of Paris and some other large universities seems determined by a specific effect of attraction rather than by any geographic factor.

5. Conclusion

By weakening topological axiomatics, pretopology provides a conceptual framework which opens the path to a wider range of applications. This concept has already been used in pattern recognition, for example, in image analysis [16], optical character recognition [15], and cluster analyses techniques [10]. In this paper we show that this concept can help formulate and treat structural analysis problems in a unified manner when connections between the elements of a population are diverse. This method has been applied in spatial economics [2] as well as in transportation economics [4,5]. It enables us to extract knowledge on the structure of a population from connections that exist between the elements. The aim is to find significant connections between groups of interdependent elements. Examples and software corresponding to this method are available at http://lass.univ-lyon1.fr/softs/index.html.

References

[1] C. Autant-Bernard, N. Massard, Scentific interactions, geographic spilovers and innovation, An empirical study on the French case, Second submit to Papers in Regional Science.

[2] Z.T. Belmandt, Manuel de prétopologie et ses applications, Hermès, Paris, 1993.

[3] G. Birkhoff, Lattice Theory, AMS Colloq Publications, Providence, RI, 1940 (3rd edition, 1967).

[4] S. Bonnevay, M. Lamure, C. Largeron, N. Nicoloyannis, A pretopology-based algorithm for structuring data in spaces endowed with weak structure, in: Ordinal Symbolic Data Analysis'98, Amherst (USA), 1998.

[5] S. Bonnevay, M. Lamure, C. Largeron, N. Nicoloyannis, A pretopological approach for structuring data in non-metric spaces, Electronic Notes in Discrete Mathematics, vol. 2 (1999). Available from http://www.elsevier.nl/gej-ng/31/29/24/32/23/show/Products/notes/index htt#002 (visited 06/28/2001).

[6] S. Bonnevay, C. Largeron, Data analysis based on minimal closed subsets, in: H.A.L. Kiers, et al. (Eds.), Data Analysis Classification and Related Methods, Springer, Berlin, 2000, pp. 303–308.

[7] E. Cech, Topological Spaces, Wiley, New York, 1966.

[8] D. Dikranjan, W. Tholen, Categorical Structure of Closure Operators, Kluwer Academic publishers, Dordrecht, 1995.

[9] M. Frechet, Espaces Abstraits, Hermann, Paris, 1928.

[10] C. Frelicot et al., in: A Pretopological Approach for Pattern Recognition with Reject Options, Lecture Notes in Computer Science, vol. 1451, 1998, pp. 707–715.

[11] F. Harari, R. Norman, D. Cartwright, Structural Models – An introduction the Theory of Directed Graphs, Wiley, New York, 1965.

[12] C. Kuratowski, Topologie, Warsaw, 1952.

[13] C. Largeron, S. Bonnevay, Une méthode de structuration par recherche de fermés minimaux – application à la modélisation de flux de migrations intervilles, Société Francophone de Classification'1997, Lyon, France, July 1997, pp. 111–118.

[14] C. Largeron, N. Massard, La géographie des collaborations scientifiques en France: une étude de la structuration des co-publications entre départements, Revue d'économie régionale et urbaine No. 1, 2001.

[15] F. Lebourgeois, H. Emptoz, Pretopological approach to the recognition: application for printed characters, CNED'96, Nantes, France, July 1996.

[16] A. Meziane et al., Satellite image segmentation by mathematical pretopology and automatic classification, Image processing, signal processing and synthetic aperture radar for remote sensing, London (England), SPIE Proc. 3217 (1997) 232–236.

[17] J. Warfeld, Societal systems, Planning, Policy and Complexity, Wiley, New York, 1976.