

# Ensuring the Drawability of Extended Euler Diagrams for up to 8 Sets

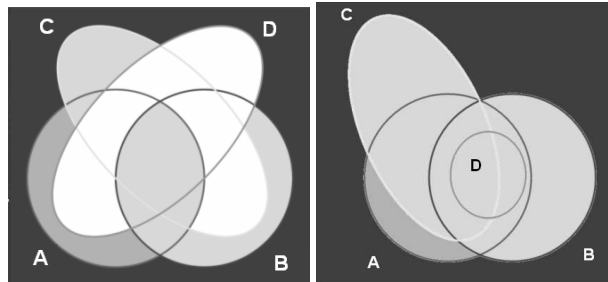
Anne Verroust<sup>1</sup> and Marie-Luce Viaud<sup>2</sup>

<sup>1</sup> INRIA Rocquencourt, B.P. 105, 78153 LE CHESNAY Cedex France. e-mail: Anne.Verroust@inria.fr

<sup>2</sup> INA, 4, avenue de l'Europe 94366 BRY SUR MARNE France. e-mail:luce@ina.fr

**Abstract.** This paper shows by a constructive method the existence of a diagrammatic representation called extended Euler diagrams for any collection of sets  $X_1, \dots, X_n$ ,  $n < 9$ . These diagrams are adapted for representing sets inclusions and intersections: each set  $X_i$  and each non empty intersection of a subcollection of  $X_1, \dots, X_n$  is represented by a unique connected region of the plane. Starting with an abstract description of the diagram, we define the dual graph  $G$  and reason with the properties of this graph to build a planar representation of the  $X_1, \dots, X_n$ . These diagrams will be used to visualize the results of a complex request on any indexed video databases. In fact, such a representation allows the user to perceive simultaneously the results of his query and the relevance of the database according to the query.

## 1 Introduction



**Fig. 1.** Venn diagram (left) and Extended Euler Diagram (right) built from the fields (A) "Paris", (B) "works", (C) "subway", (D) "beltway". Grey levels of the regions are linked to the number of documents in the region : white regions are empty ones

Nowadays, enhancing the visualization of the results of complex queries in large databases is becoming a challenging and useful task [3, 2]. We propose to tackle this problem with providing the user a semantic cartography of the set of documents answering his complex query. Let us illustrate our approach with a complex query built by students on INA's database of video documents. They had to work on the evolution of transportation in Paris. Figure 1 shows two diagrams built from the fields *Paris*, *works*, *subway* and *beltway*. The leftmost diagram is a Venn diagram which contains all combinations of the fields, the rightmost one is an *Extended Euler Diagram* where only

non-empty regions appear. These two diagrams show the repartition of the documents in the database according to users criteria, given the contribution of each field in order to help the user to elaborate a new formulation of the query. Moreover, these diagrams may be used as an interface to compose interactively boolean expressions by selecting regions, helping users not familiar with brackets and operator order to build complex queries.

This work has been done in the context of INA: INA is the French radio and television legal deposit center since the law of June 1992. INA's archives contain more than 3 million documents representing approximately 400 000 hours of video and 500 000 hours of audio programs. Recently the consultation of the database has been opened to non-professionals of documentation such as researchers or producers, creating the need of more convivial interfaces. When a new document is inserted in the database, it is described and analyzed by professionals, using predefined structured classifications (thesaurus, controlled lists...) to allow, as much as possible, non ambiguous identifications of the documents. As a consequence, many indices of the controlled structures are exclusive and, in the associated diagram representation, many regions may be empty. This observation enhances the relevance of Extended Euler Diagrams (EED) compared to Venn diagrams. Indeed, if the intersection of  $k$  fields is empty, any intersection of those  $k$  fields with other ones is empty too. Then EED may contain significantly fewer regions than Venn diagrams, leading to a much more readable cartography. However, in such an application, we need to have an Extended Euler Diagram for any combination of fields and a method to build it dynamically.

This paper is organized as follows:

- We first describe previous works on diagrams and introduce a description of **Extended Euler Diagrams** (EED in the following), and their properties. Then, we give a formal definition of EED and introduce a dual representation, in terms of graph formalization: the  $L$ -connected labelled graphs.
- In the second section, we show that the drawability of the EED is related to the planarity of the  $L$ -connected labelled graph and give limitations on the number of sets being represented by EED.
- The third section contains a sketch of the proof which builds, for any collection of  $n < 9$  sets, a planar  $L$ -connected labelled graph representing it. We first build a minimal subgraph on a specific subset of vertices, then we show that this graph is planar and that any other vertex can be inserted in this minimal subgraph without breaking the planarity.

## 2 EED definition

### 2.1 Previous works

Given  $X = \{X_1, X_2, \dots, X_k\}$  a collection of non empty distinct sets, we want to build a graphic representation which shows information about the sets and their intersections on a plane. Let  $Y = \{Y_1, Y_2, \dots, Y_{2^k}\}$  be the collection of all intersections between the  $X_i$  and  $Y_r = \{Y_i | Y_i \in Y \text{ and } Y_i \neq \emptyset\}$ . Euler diagrams [4] could be used but appear to be too restrictive for our purpose. In fact, an Euler diagram consists of a collection of simple closed and convex curves, called contours, which split the plane into zones.

Each set  $X_i$  is associated with an unique contour, and is represented by the interior of this contour. However, because of the convexity constraint, some  $Y_r$  cannot be drawn with an Euler diagram when the number of sets  $X_i$  is equal to 4 (for a discussion on Euler diagrams restrictions, the reader may consult [9, 11]).

The concrete Euler diagrams proposed in Flower and Howse's approach [5], used in another purpose [6], are very well defined but are still very restrictive. In fact, concrete Euler diagrams are Euler diagrams with still very strong constraints. The first constraints introduced at the curve level, make hypothesis on the set of intersections being drawn: each segment of curve delimits the interior and the exterior of exactly one set, and each intersection of curves is the crossing of exactly two contours. The introduction of "exactly" is very useful to specify formally the problem and its dual formulation with graphs, but eliminates the numerous cases in which the set of subsets built from the intersections of the  $X_i$  does not have such properties.

According to our purposes, we propose an extension of Euler diagrams which makes drawable any collection  $Y_r$  from a set  $X$  of  $X_i$  such that  $card(X) \leq 8$ .

Such diagrams are characterized by the following properties:

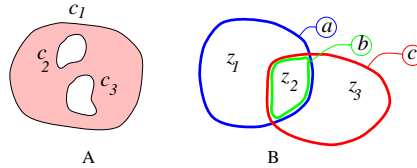
- An intersection point may intersect more than two contours,
- A curve segment may be part of more than one contour,
- Each non empty  $Y_i$  is associated with a unique zone,
- Each set  $X_i$  is associated with a set of zones whose union forms a connected planar region. This region may not be convex and may contain holes.

## 2.2 Extended Euler Diagrams

**Definition 1.** Let  $L$  be a finite set of labels and  $C$  a set of simple closed (Jordan) curves in  $\mathbb{R}^2$ . We say that  $C$  is labelled by  $L$  when each curve  $c$  of  $C$  is associated with a couple  $(\lambda(c), sign(c))$  where  $\lambda(c) \in L$  and  $sign(c) \in \{+, -\}$ .

To each labelled curve  $c$  of  $C$  corresponds a zone  $\zeta(c)$  defined by:

- if  $sign(c) = +$ , then  $\zeta(c) = int(c)$
- if  $sign(c) = -$ , then  $\zeta(c) = ext(c)$



**Fig. 2.** A: a zone with two holes  $sign(c_2) = sign(c_3) = -$ ; B: an extended Euler diagram with  $m(z_1) = \{a\}$ ,  $m(z_2) = \{a, b, c\}$  and  $m(z_3) = \{c\}$ .

**Definition 2.** An extended Euler diagram is a triple  $(L, C, Z)$  whose components are defined as follows:

1.  $L$  is a finite set of labels
2.  $C$  is a set of Jordan curves labelled by  $L$  and verifying:
  - (a)  $\forall l \in L, \exists c \in C, \lambda(c) = l$  and  $sign(c) = +$ .
  - (b) if  $\lambda(c) = \lambda(c')$ ,  $c \neq c'$  and  $sign(c) = sign(c')$  then  $c$  and  $c'$  do not intersect

- (c) if  $\lambda(c) = \lambda(c')$ ,  $c \neq c'$  and  $\text{sign}(c) = +$ , then  $\text{sign}(c') = -$  and  $c' \subset \text{int}(c)$
3.  $Z$  is a set of zones corresponding to the planar partition defined by  $C$ .  
Each zone  $z$  of  $Z$  is associated with a set of labels  $m(z)$  defined by
- (a)  $m(z) = \{l \in L \mid \forall c \in C, \text{ if } \lambda(c) = l \text{ then } z \subset \zeta(c)\}$   
(b) if  $m(z) = m(z')$  and  $m(z) \neq \emptyset$ , then  $z = z'$
- We note  $Z_\emptyset$  the set of zones associated with an empty set of labels.  
 $Z_\emptyset$  contains at least the zone  $z_\emptyset = \bigcap_{\{c \mid \text{sign}(c) = +\}} \text{ext}(c)$ .

The set of extended Euler diagrams is noted  $\mathcal{EED}$ .

As a matter of fact, we have introduced Jordan curves to define zones, but those notions are equivalent. In the following, we will use rather the zones formalization.

**Definition 3.** Let  $X = \{X_1, X_2, \dots, X_k\}$  be a set of non empty distinct subsets of  $\mathcal{X}$ ,  $Y_r = \{Y_1, Y_2, \dots, Y_m\}$  the set of all possible non empty intersections between the  $X_i$  ( $m \leq 2^k$ ). We say that the extended Euler diagram  $(L, C, Z)$  is a diagram representation of  $X$  if and only if:

1. there is a bijection  $\psi : L \rightarrow X; l \mapsto x$
2.  $\phi : Z \setminus Z_\emptyset \rightarrow Y_r; z \mapsto y$  defined by  $\phi(z) = y = \bigcap_{l \in m(z)} \psi(l)$  is a bijection.

### 2.3 L\_connected labelled graphs

A  $L$ -connected labelled graph is a labelled graph which ensures that, for any label  $l$  of  $L$ , there is a path connecting all the vertices labelled by  $l$ . We give in this section a more formal definition of  $L$ -connected labelled graphs.

**Definition 4.** A labelled graph is a triple  $G(L, V, E)$  where:

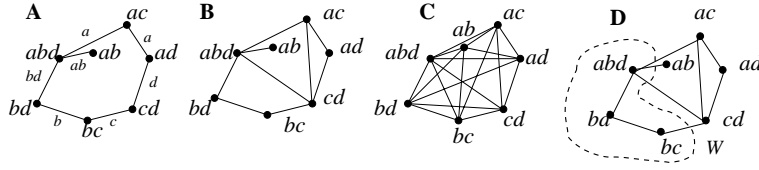
1.  $L$  is a finite set of labels
2.  $V$  is a set of labelled vertices, i.e.:
  - (a) each vertex  $v$  is labelled with a set of labels  $m(v) \subseteq L$
  - (b) two distinct vertices  $v$  and  $w$  of  $V$  have distinct sets of labels.
3.  $E$  is a set of edges such that:
  - (a) each edge  $e = (v, w)$  of  $E$  is labelled with a set of labels  $m(e) = m(v) \cap m(w)$
  - (b) if  $e \in E$  then  $m(e) \neq \emptyset$

In the rest of the paper,  $L(W)$  will be the set of labels associated with the vertices of  $W$ , i.e.  $L(W) = \bigcup_{v \in W} m(v)$ , where  $W$  is a set of labelled vertices.

**Definition 5.** Let  $G(L, V, E)$  be a labelled graph (cf. figure 3 for an example).

- Let  $l$  be a label of  $L$ . We say that  $G(L, V, E)$  is  $l$ -connected if and only if the subgraph  $G'$  of  $G(L, V, E)$  on the set  $V'$  of vertices of  $V$  having  $l$  in its set of labels is connected.
- $G(L, V, E)$  is said  $L$ -connected if and only if it is  $l$ -connected for all  $l$  in  $L$ .
- $G(L, V, E)$  is said  $L$ -complete when  $E$  is defined by:  
 $E = \{(v, w) \mid v \in V, w \in V \text{ and } m(v) \cap m(w) \neq \emptyset\}$
- A vertex  $v$  of  $V$  is said  $L$ -connectable to a subset  $W$  of  $V$  if and only if  $m(v) \subseteq L(W)$ .

In fact, given  $L$  and  $V$ , there exists only one  $L$ -complete labelled graph  $G(L, V, E)$ , noted  $G(L, V, E_c)$  and any  $L$ -connected labelled graph  $G(L, V, E)$  is a subgraph of  $G(L, V, E_c)$ .



**Fig. 3.**  $L = \{a, b, c, d\}$ ,  $V = \{abd, bd, bc, cd, ad, ac\}$ ,  
 $E_1 = \{(abd, bd), (bd, bc), (bc, cd), (cd, ad), (ad, ac), (ac, abd)\}$ ,  
 $E = E_1 \cup \{(abd, cd), (cd, ac)\}$ ,  $E_c = E \cup \{(ac, bc), (bc, abd), (abd, ad), (ad, bd), (bd, cd)\}$ .  
A:  $G(L, V, E_1)$  is a<sub>L</sub>-connected, b<sub>L</sub>-connected but it is not c<sub>L</sub>-connected and d<sub>L</sub>-connected;  
B:  $G(L, V, E)$  is a L<sub>L</sub>-connected labelled graph;  
C:  $G(L, V, E_c)$  is the corresponding L<sub>L</sub>-complete graph;  
D: ac is L<sub>L</sub>-connectable to  $W = \{abd, bd, bc\}$ .

**Definition 6.** Let  $Y_r = \{Y_1, Y_2, \dots, Y_m\}$  be the subset of  $Y$  which elements are the non empty intersections between the  $X_i (m \leq 2^k)$  We say that the L<sub>L</sub>-connected labelled graph  $G(L, V, E)$  is a graph representation of  $X$  if and only if:

1. there is a bijection  $\lambda : L \rightarrow X = \{X_1, \dots, X_k\}$ ;  $l \mapsto x$
2.  $\chi : V \rightarrow Y_r$ ;  $v \mapsto y$  defined by  $\chi(v) = y = \bigcap_{l \in m(z)} \lambda(l)$  is a bijection.

In the following, we note:

- $\mathcal{G}^V(L, V)$  the set of L<sub>L</sub>-connected labelled graphs associated with a given set of label  $L$  and a set  $V$  of labelled vertices.
- $\mathcal{G}_P^V(L, V)$  the set of planar graphs belonging to  $\mathcal{G}^V(L, V)$ .

### 3 Drawability of EED

We have defined extended Euler diagrams and L<sub>L</sub>-connected labelled graphs. These two notions are related:

- a planar L<sub>L</sub>-connected graph is deduced by duality from an extended Euler diagram (cf. section 3.1).
- the drawing of a planar L<sub>L</sub>-connected graph leads to a class of EED. We describe briefly in section 3.2 a method to build one of them.

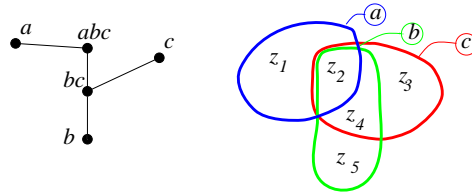
Thus it is equivalent to show the drawability of an extended Euler diagram and the planarity of the L<sub>L</sub>-connected labelled graph associated with it. This observation leads us to study the planarity of L<sub>L</sub>-connected labelled graphs with respect to the cardinality of the set of labels  $L$ .

#### 3.1 From extended Euler diagrams to planar L<sub>L</sub>-connected labelled graphs

**Definition 7.** The mapping  $dual : \mathcal{EED} \rightarrow \mathcal{G}; (L', C, Z) \mapsto G(L, V, E)$  is defined by:  $G(L, V, E) = dual((L', C, Z))$  if and only if

- (i) there is a one to one mapping between  $L'$  and  $L$
- (ii) there is a bijection  $\delta : Z \rightarrow V$ ;  $z \mapsto v$  such that  $m(z) = m(\delta(z))$
- (iii)  $e = (v, w) \in E$  if and only if  $\delta^{-1}(v)$  and  $\delta^{-1}(w)$  are adjacent along a portion of curve of non null length in the planar partition formed by  $C$ .

As a consequence, if  $(L, C, Z)$  is a diagram representation of  $X$ , then  $dual((L, C, Z))$  is a graph representation of  $X$ .

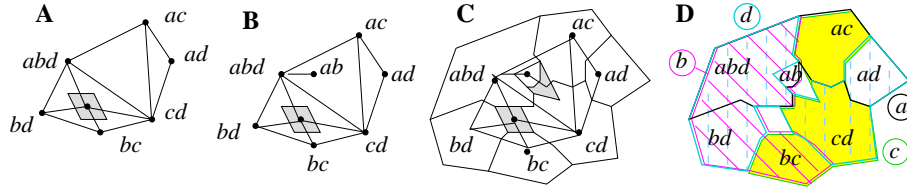


**Fig. 4.** An extended Euler diagram  $(L, C, Z)$  and its dual. We have  $m(a) = m(z_1) = \{a\}$ ,  $m(abc) = m(z_2) = \{a, b, c\}$ ,  $m(c) = m(z_3) = \{c\}$ ,  $m(bc) = m(z_4) = \{b, c\}$  and  $m(b) = m(z_5) = \{b\}$ .

### 3.2 From L\_connected labelled graphs to extended Euler diagrams

**Proposition 1.** *If there is a planar L\_connected graph  $G = G(L, V, E)$  representing  $X$ , then there is a class of extended Euler diagrams  $(L, C, Z)$  representing  $X$ . These diagrams are such that  $G(L, V, E) = \text{dual}((L, C, Z))$ .*

Let us give the outline of the building of one Extended Euler diagram from a planar L\_connected labelled graph with the example of figure 5. Following [1], we generate a straight-line drawing  $D(G)$  from  $G$ . When a face is triangular, it is obvious to draw the contours of the zones in this face. In other cases, we extend  $G$  by introducing special edges and vertices to obtain a triangulated graph  $G'$ . These vertices correspond to zones having an empty set of labels and provide a better control on the drawing of the resulting diagram. The dangling edges deserve a slightly different process as shown in figure 5.



**Fig. 5.** The construction of an extended Euler diagram from a drawing of a the planar L\_connected graph of figure 3. Each internal empty zone is drawn in grey and is associated with a new vertex. A: a drawing of a L\_connected labelled graph without the dangling edge  $(abd, ab)$  and the empty zone associated with a non triangular face. B: the dangling edge  $(abd, ab)$  is drawn. C: the graph and its associated zones. D: the extended Euler diagram with internal zones associated with an empty set of labels.

$$\begin{array}{ccc}
 G(L, V, E) & \xrightarrow{\text{draw}} & D(G) & \xrightarrow{\text{extend}} & D(G') \\
 & \swarrow \text{dual} & \downarrow & \searrow \text{diag} & \\
 & & (L, C, Z) & & 
 \end{array}$$

### 3.3 Planarity of L\_connected labelled graphs

To study the planarity of a L\_connected labelled graph, we use the graphs  $K_n$  and  $K_{n,n}$ ,  $n \geq 2$

-  $K_n$  is the complete graph defined on  $n$  vertices: in  $K_n$ , every vertex is adjacent to

every other vertex

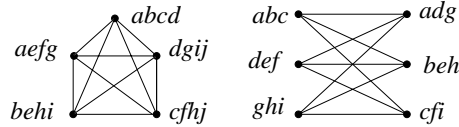
-  $K_{n,n}$  is the complete bipartite graph consisting of two disjoint vertex sets  $V = \{v_1, \dots, v_n\}$ ,  $W = \{w_1, \dots, w_n\}$  and the edge set  $E = \{v_i, w_j | 1 \leq i, j \leq n\}$  and Kuratowski's characterization of planar graphs [8]:

**Theorem 1.** *A graph is planar if and only if it does not contain a subdivision of  $K_5$  or  $K_{3,3}$  as a subgraph.*

We already know that, if  $\text{card}(V) = 2^{\text{card}(L)}$  the diagram to draw is a Venn diagram which has a planar representation (cf. [10]) for any value of  $\text{card}(L)$ . But this property does not hold in the general case. In fact we have<sup>3</sup>:

**Proposition 2.** *Let  $k$  be the cardinality of  $L$ . When  $k \geq 9$ , there exists at least a set of labelled vertices  $V$  for which all the graphs of  $\mathcal{G}^V(L, V)$  are non planar.*

*Proof.* Suppose  $L = \{a, b, c, d, e, f, g, h, i\}$  and  $V = \{abc, def, ghi, adg, beh, cfi\}$ . Then  $\mathcal{G}^V(L, V)$  contains only one L-connected labelled graph which is a  $K_{3,3}$  (cf. figure 6).  $\square$



**Fig. 6.** Two non planar L-connected labelled graphs: on the left a  $K_5$  and on the right a  $K_{3,3}$

In the next section, we will show that for  $\text{card}(L) < 9$  and for any set of labelled vertices  $V$  on  $L$ ,  $\mathcal{G}^V(L, V)$  contains at least one planar graph.

## 4 The constructive proof

### 4.1 Sketch of the constructive proof

In the rest of the paper,  $L$  will denote a set of labels and  $V$  a set of labelled vertices on  $L$ . We add constraints on  $V$  using the following results:

**Definition 8.** *Let  $V$  be a set of labelled vertices and  $v$  and  $w$  two vertices of  $V$ ,*

-  *$v$  and  $w$  are said label\_disjoint when  $m(v) \cap m(w) = \emptyset$*

-  *$v$  is said label\_included in  $w$  when  $m(v) \subset m(w)$ .*

**Proposition 3.** *Let  $W$  be a set of vertices such that  $W \subset V$ , every vertex of  $W$  is label\_included in a vertex of  $V$  and  $V_r = V \setminus W$ . Then, if  $\mathcal{G}^V(L, V_r)$  contains a planar graph  $G(L, V_r, E_r)$ ,  $\mathcal{G}_P^V(L, V)$  is not empty.*

*Proof.* Let  $w \in W$  and  $v \in V$  be such that  $w$  is label\_included in  $v$ . Then if we add the edge  $e = (v, w)$  to  $G(L, V_r, E_r)$ , we obtain a L-connected graph  $G(L, V_r \cup \{w\}, E_r \cup \{e\})$  which is still planar (the addition of  $e$  cannot contribute to add a  $K_{3,3}$  or a  $K_5$  in  $(L, V_r \cup \{w\}, E_r \cup \{e\})$ ). By augmenting the same way the graph  $G(L, V_r, E_r)$  for each  $w \in W$ , we obtain at the end a L-connected planar graph on  $V$ .  $\square$

<sup>3</sup> The following proposition is another version of the planarity results for Euler's Circles presented by Lemon and Pratt in [9].

**Corollary 1.** *Let  $W$  be the subset of  $V$  formed by vertices associated with only one label and  $V_r = V \setminus W$ . Then, if  $\mathcal{G}^{\mathcal{V}}(L, V_r)$  contains a planar graph  $G(L, V_r, E_r)$ ,  $\mathcal{G}_P^{\mathcal{V}}(L, V)$  is not empty.*

*Proof.* We use the fact that a vertex  $w$  of  $W$  is either label\_included in a vertex  $v$  of  $V$  or label\_disjoint of any vertex  $v$  of  $V$ .  $\square$

Using the previous corollary, we will restrict ourselves to sets of labelled vertices  $V$  on  $L$  satisfying:

(H1)  $\forall v, w \in V$ , if  $v$  is label\_included in  $w$  then  $v = w$ .

(H2) any vertex  $v$  of  $V$  has more than one label in  $m(v)$ .

For a set  $V$  of labelled vertices satisfying (H1) and (H2), we proceed as follows to show that  $\mathcal{G}_P^{\mathcal{V}}(L, V)$  is not empty when  $\text{card}(L) < 9$ :

1. we choose a subset  $V_0$  of vertices of  $V$  among those satisfying  $L(V_0) = L(V)$  and build a  $L$ -connected planar graph  $G_0(L, V_0, E_0)$  on  $V_0$  (cf section 4.2).
2. we build a partition of  $V = V_0 \cup V_1 \cup \dots \cup V_k$ ,  $k \leq \text{card}(V_0)$  (cf. section 4.3).
3. then, for each  $V_i$ ,  $i > 1$ , we show how to extend  $G_0(L, V_0 \cup V_1 \dots \cup V_{i-1}, E_{i-1})$  to obtain a  $L$ -connected planar graph  $G(L, V_0 \cup V_1 \dots \cup V_i, E_i)$ . This is the subject of section 4.4

## 4.2 Choice of $V_0$ and construction of $G_0$

Let  $\mathcal{T}(V)$  be the set of subsets  $T$  of  $V$  such that  $L(T) = L(V)$  and  $\mathcal{T}_0(V)$  be the subset of  $\mathcal{T}(V)$  formed by the elements  $T$  of  $\mathcal{T}(V)$  having a minimum number of vertices. Given  $T$  and  $T'$  in  $\mathcal{T}_0(V)$ , we rename the vertices of  $T$  and  $T'$  w.r.t. the cardinality of their associated sets of labels, i.e.  $T = \{v_0, v_1 \dots v_p\}$  and  $T' = \{v'_0, v'_1 \dots v'_p\}$  with  $\text{card}(m(v_i)) \geq \text{card}(m(v_j))$  and  $\text{card}(m(v'_i)) \geq \text{card}(m(v'_j))$  when  $i \leq j$ .

We say that  $T \geq_L T'$  if and only if:

$\exists k \leq p, \forall i < k, \text{card}(m(v_i)) = \text{card}(m(v'_i))$  and  $\text{card}(m(v_k)) \geq \text{card}(m(v'_k))$ .

$\mathcal{T}_{\max}(V)$  is the set composed by the maximal elements of  $\mathcal{T}_0(V)$  for  $\geq_L$ .

In the rest of the paper,  $V_0$  denotes an element of  $\mathcal{T}_{\max}(V)$ . Moreover, for any subset  $W$  of  $V$ ,  $Lu(W)$  denotes the set of labels belonging to only one  $m(v)$  for  $v \in W$  and  $Lu_W(v) = Lu(W) \cap m(v)$ .

The definition of  $\mathcal{T}_{\max}(V)$  and the hypothesis (H1) and (H2) on  $V$  imply the following results:

-1- for any vertex  $v$  of  $V_0$ ,  $Lu_{V_0}(v)$  contains at least one label, thus  $\text{card}(Lu(V_0)) \geq \text{card}(V_0)$ . As any  $m(v)$  contains at least two labels, we have:  $\text{card}(V_0) < \text{card}(L)$

-2- Any subset  $V'$  of  $V_0$  inherits the minimality properties from  $V_0$ :  $V'$  is such that we cannot have neither

(a)  $\exists V'' \subseteq V$ , s.t.  $L(V') \subseteq L(V'')$  and  $\text{card}(V') > \text{card}(V'')$

nor

(b)  $\exists V'' \subseteq V$ , s.t.  $L(V') \subseteq L(V'')$  and  $\text{card}(V') = \text{card}(V'')$  and  $V' <_L V''$

These two points will be used when computing a  $L$ -connected planar graph on  $V$ :

Let us suppose that  $V_0 = \{v_1, v_2 \dots v_k\}$  and  $\text{card}(L) = n$ . Then using (1), we have  $\text{card}(Lu(V_0)) \geq k$ . Moreover, because of (2),  $V \setminus V_0$  cannot contain a vertex  $v$  such that  $Lu_{\{V_0\}}(v_i) \cup Lu_{\{V_0\}}(v_j) \subset m(v)$  where  $v_i$  and  $v_j$  are two distinct vertices of

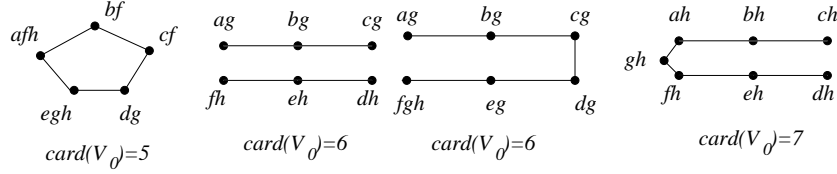


$V_0$ . In fact, in such a case,  $V' = V_0 \cup \{v\} \setminus \{v_i, v_j\}$  would contradict the hypothesis of minimality of  $V_0$ .

Then, the presence of labels of  $Lu(V_0)$  in the set of labels of a vertex of  $V \setminus V_0$  is strongly constrained, and this fact reduces drastically the number of cases to consider at each step of the proof.

**Proposition 4.** *If  $\text{card}(L) < 9$ ,  $\mathcal{G}_P^V(L, V_0)$  is not empty. We note  $G_0$  an element of  $\mathcal{G}_P^V(L, V_0)$ .*

*Proof.* Let  $G(L, V_0, E_0)$  be a  $L$ -connected graph of  $\mathcal{G}^V(L, V_0)$  having a minimal number of edges (we give examples in figure 7). Then, as  $\text{card}(L) < 9$  and  $\text{card}(V_0) < \text{card}(L)$ , we have the following cases to consider:



**Fig. 7.** Examples of labelled graphs  $G_0$

–  $\text{card}(V_0) \leq 4$ .  $\mathcal{G}^V(L, V_0) = \mathcal{G}_P^V(L, V_0)$  ( $K_5$  and  $K_{3,3}$  have respectively 5 and 6 vertices).

–  $\text{card}(V_0) = 5$ . If  $G(L, V_0, E_0) = K_5$ , as  $E_0$  is minimal, each edge of  $K_5$  would be associated with a label, thus  $\text{card}(L) \geq 10$ , which contradicts the hypothesis.

–  $\text{card}(V_0) = 6$ . We have  $\text{card}(Lu(V_0)) \geq 6$  and  $\text{card}(L) \leq 8$ . Then  $L$  contains one or two labels which do not belong to  $Lu(V_0)$ .

If  $L \setminus Lu(V_0) = \{l\}$ , then  $E_0$  consists in a path joining the vertices having  $l$  in their set of labels and  $G(L, V_0, E_0)$  is planar.

If  $L \setminus Lu(V_0) = \{l, l'\}$ : to build a  $L$ -connected graph on six vertices connecting two labels, we need less than 10 edges. Such a graph cannot contain any  $K_{3,3}$  or  $K_5$ .

–  $\text{card}(V_0) = 7$ . As  $L \setminus Lu(V_0) = \{l\}$ ,  $E_0$  consists in a path joining the vertices having  $l$  in their set of labels.  $\square$

### 4.3 Construction of a partition of $V \setminus V_0$

Let  $G_0 = (L, V_0, E_0)$  be a planar  $L$ -connected graph. To extend  $G_0$  with the vertices of  $V \setminus V_0$ , we first build a partition  $V \setminus V_0$  into a family of sets  $V_1, \dots, V_k$  with  $k \leq \text{card}(L)$ : A vertex  $v$  of  $V \setminus V_0$  belongs to  $V_i$  if and only if the addition of  $i$  edges connecting  $v$  to  $V_0$  is necessary and sufficient to  $L$ -connect  $v$  to  $V_0$ . We note  $\mathcal{W}_{MLN}(v, V_0)$  the set of subsets  $W$  of  $V_0$  such that  $\text{card}(W) = i$  and  $v$  is  $L$ -connectable to  $W$ .

One shall notice that hypothesis (H1) on  $V$  implies that  $V_1$  is empty.

Before extending  $G_0$  with the  $V_i$ , we will give general results on the  $V_{i, i>0}$ .

**Lemma 1.** *If  $v \in V_n$  and if  $W_n = \{w_1, \dots, w_n\}$  is an element of  $\mathcal{W}_{MLN}(v, V_0)$ , then  $\text{card}(m(v)) \geq n$  and  $m(v) = \{l_1..l_n\} \cup L_r$ , with  $L_r \subset L(W_n)$  and  $l_i \in Lu_{W_n}(w_i)$ .*

*Proof.* As  $v$  is in  $V_n$ ,  $v$  is L\_connectable to  $W_n$ . Therefore, if there was  $w_i$  in  $W_n$  such that  $m(v) \cap Lu_{W_n}(w_i) = \emptyset$ , then  $v$  would be L\_connectable to  $W_n \setminus \{w_i\}$  and  $W_n \notin \mathcal{W}_{\mathcal{MIN}}(v, V_0)$ .  $\square$

**Proposition 5.**

If  $V_{n,n \geq 2}$  is not empty then  $\forall v \in V_n, \forall W_n \in \mathcal{W}_{\mathcal{MIN}}(v, V_0), n \leq \frac{\text{card}(L(W_n))}{2}$

*Proof.* Sketch of proof (the detailed proof can be found in [12]): Let us suppose that  $v \in V_n$  and  $W_n \in \mathcal{W}_{\mathcal{MIN}}(v, V_0)$ . Then using lemma 1, we show that if  $L(W_n) < 2n$ , one can find a set of vertices  $W' \subset W \cup \{v\}$  such that  $L(W) = L(W')$  and either  $\text{card}(W') < \text{card}(W)$  or  $\text{card}(W') = \text{card}(W)$  and  $W <_L W'$ .  $\square$

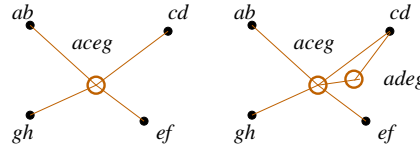
Thus, using this result, we know that  $V_n$  is empty when  $2n > \text{card}(L)$ . In particular, when  $\text{card}(L) < 9, V = V_0 \cup V_2 \cup V_3 \cup V_4$ .

**4.4 Construction of a L\_connected planar graph when  $\text{card}(L) < 9$**

For the readability of the paper, we will give here an idea of the construction. The detailed description of the construction is presented in [12].

By definition of  $V_i$ , we can add  $i$  edges connecting a vertex  $v$  of  $V_i$  to  $V_0$  and obtain a L\_connected labelled graph  $G'$ . However, our goal here is to keep also the planarity of the graph while extending it. Thus, to built a planar labelled graph on  $V$ , we insert incrementally vertices of  $V \setminus V_0$  on the L\_connected planar graph  $G_0$ . We first insert vertices of  $V_n$  with  $n$  maximal.

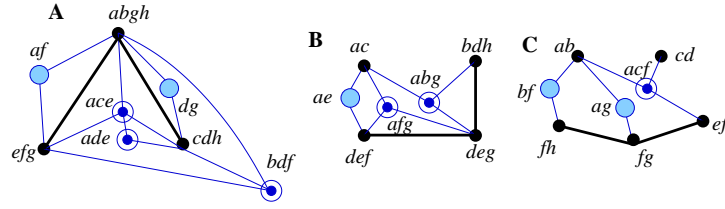
-  $V_4$  is non empty only when  $\text{card}(L) = 8$  and  $\text{card}(V_0) = 4$ . In this case,  $V_4$  contains at most 2 elements (cf figure 8): four edges are added for the first element of  $V_4$ , and two edges are necessary to connect the second element of  $V_4$ .



**Fig. 8.** Insertion of vertices of  $V_4$ .  $V = \{ab, cd, ef, gh, aceg, adeg\}, V_0 = \{ab, cd, ef, gh\}, V_4 = \{aceg, adeg\}, E_0$  is empty. In a first time,  $aceg$  is inserted by adding four edges  $(aceg, ab), (aceg, cd), (aceg, ef), (aceg, gh)$ . Then  $adeg$  is inserted by adding two edges:  $(aceg, adeg), (adeg, cd)$ .

- When it is not possible to add  $i$  edges to  $G_0$  to insert a vertex  $v$  of  $V_i$  without breaking the planarity then  $v$  is connected with another vertex of  $V_i$  already inserted in  $G_0$ , as it is the case in figure 9 A. This is always possible when  $\text{card}(L) < 9$ : otherwise this leads to a contradiction on the hypothesis on  $V_0$ . In fact, by using the partition of  $V$  in  $V_0, \dots, V_n$  in this process, we have restricted the number of cases to consider to a few generic cases when  $\text{card}(L) < 9$ .

We then obtain the following result:



**Fig. 9.** Insertion of vertices of  $V_2$  and  $V_3$  in  $G_0(L, V_0, E_0)$ .

A:  $V_0 = \{abgh, efg, cdh\}$ ,  $E_0 = \{(abgh, efg), (abgh, cdh)\}$ ,  $V_2 = \{af, dg\}$ ,  $V_3 = \{ace, ade, bdf\}$ . The vertices of  $V_3$  are inserted incrementally: the two label\_disjoint vertices  $ace$  and  $bdf$  are inserted in the two faces defined by  $abgh, efg$  and  $cdh$ . Then,  $ade$  is inserted by adding only two edges connecting  $ade$  with  $ace$  and  $cd$ .

B:  $V_0 = \{ac, def, deg, bdh\}$ ,  $E_0 = \{(def, deg), (deg, bdh)\}$ ,  $V_2 = \{ae\}$ ,  $V_3 = \{afg, abg\}$ .

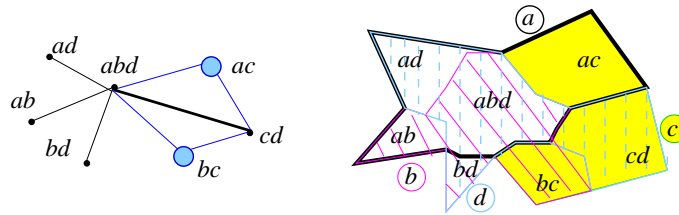
C:  $V_0 = \{ab, cd, ef, fg, fh\}$ ,  $E_0 = \{(fh, fg), (fg, ef)\}$ ,  $V_2 = \{bf, ag\}$ ,  $V_3 = \{acf\}$ .

**Theorem 2.** When  $\text{card}(L) < 9$  then for any set  $V$  of labelled vertices on  $L$ ,  $\mathcal{G}_P^V(L, V)$  is not empty.

Then, using proposition 1, we have:

**Corollary 2.** For any set of non empty distinct sets  $X = \{X_1, \dots, X_k\}$  such that  $k < 9$  there is an extended Euler diagram representing  $X$ .

*Remark 1.* Let us consider the set of vertices  $V$  of figure 5.  $V_0 = \{abd, cd\}$ ,  $V_1 = \{ab, ad, bd\}$  and  $V_2 = \{bc, ac\}$ .  $G(L, V_0, E)$  has only one edge  $(abd, cd)$  and the  $L$ -connected labelled graph built by inserting successively the vertices of  $V_2$  and  $V_1$  leads to an extended Euler diagram where each label corresponds to a connected region (cf. figure 10).



**Fig. 10.** The  $L$ -connected labelled graph built by computing the  $V_i$  and the corresponding extended Euler diagram

#### 4.5 Hypergraph vertex-planarity: an equivalent formulation of the problem

Extended Euler diagrams can be related with Johnson and Pollak's notion of planarity for hypergraphs [7]:

Let  $H = (V, E)$  be an hypergraph and  $X = \{X_1, \dots, X_k\}$  be a set of non empty distinct subsets of  $\mathcal{X}$  such that there are:

- a one-one map  $\epsilon$  from the set of hyperedges  $E$  and  $X = \{X_1, \dots, X_k\}$ ,
- a map  $\sigma$  between  $V$  and the set of all possible non empty intersections between the

$X_i, Y_r = \{Y_1, Y_2, \dots, Y_m\}$  satisfying:  $\forall v \in V, v$  belongs to the hyperedge  $e$  of  $E$  if and only if  $\sigma(v) \subseteq \epsilon(e)$ .

If an extended Euler diagram  $(L, C, Z)$  is a diagram representation of  $X$ , then  $(L, C, Z)$  is a vertex-based diagram representing the hypergraph  $H = (V, E)$  and  $H$  is vertex-planar according to Johnson and Pollak's definition.

Interpreting theorem 2, we obtain the following result on hypergraphs:

**Corollary 3.** *Any hypergraph having at most eight hyperedges is vertex-planar.*

## 5 Conclusion

We have shown that there exists a planar  $L$ -connected graph for any collection of intersections between up to eight sets  $\{X_1, \dots, X_k\}$ . This planar  $L$ -connected graph can be used to build an extended Euler diagram representing  $\{X_1, \dots, X_k\}$ .

Interpreting our work using Johnson and Pollak's notion of planarity [7] we have shown in this paper that any hypergraph having at most eight hyperedges is vertex-planar.

We are currently working on the algorithm to produce the planar graph and the extended Euler diagram. However, to reach the purposes described in the introduction, i.e. to create a semantically structured map of the results of a complex query, we have to address a few more tasks. Indeed, for most of the collections of intersections, there exists many planar graphs satisfying the constraint of  $L$ -connectivity, and the graph built from the proof may not be the most adapted to our purposes. Then at this graph level, we may have to introduce some graphical criterion to provide the user the most readable diagram. Moreover, we still have to find the best embedding according to visibility and usability criterion.

## References

1. G. Di Battista, P. Eades, R. Tamassia, and I.G. Tollis. *Graph Drawing: algorithms for the visualization of graphs*. Prentice Hall, 1999.
2. T. Catarci, M.F. Costabile, S. Levialdi, and C. Batini. Visual query systems for databases: a survey. *Journal of Visual Languages and Computing*, 8:215–260, 1997.
3. M. Consens. *Creating and filtering structural data visualizations using Hygraph patterns*. Tech. Rep. CSRI–302, Toronto, Canada, February 1994.
4. L. Euler. *Lettres a une Princesse d'Allemagne sur divers sujets de physique et de philosophie*, volume T. 2, letters 102-108. Berne, Socit Typographique, 1775.
5. J. Flower and J. Howse. Generating Euler diagrams. In *Diagrams 2002*, pages 61–75, Callaway Gardens, Georgia, USA, April 2002.
6. J. Gil, S. Kent, and J. Howse. Formalizing spider diagrams. In *IEEE Symposium on Visual Languages*, pages 130–137, Tokyo, September 1999. IEEE Computer Society Press.
7. D.S. Johnson and H. O. Pollak. Hypergraph planarity and the complexity of drawing Venn diagrams. *Journal of Graph Theory*, 11(3):309–325, 1987.
8. K. Kuratowski. Sur le probleme des courbes gauches en topologie. *Fundamenta Mathematicae*, 15:271–283, 1930.
9. O. Lemon and I. Pratt. Spatial logic and the complexity of diagrammatic reasoning. *Machine Graphics and Vision (Special Issue on Diagrammatic Reasoning)*, 6(1):89–109, 1997.

10. F. Ruskey. A survey of Venn diagrams. *The electronic journal of combinatorics*, March 2001. <http://www.combinatorics.org/Surveys/ds5/VennEJC.html>.
11. S-J. Shin and O. Lemon. Diagrams. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. 2003. <http://plato.stanford.edu/entries/diagrams/>.
12. A. Verroust and M-L. Viaud. Ensuring the drawability of extended Euler diagrams for up to 8 sets. Technical Report RR-4973, INRIA, October 2003.