TOPIC MAPS ET HYPERGRAPHES

Pascal Auillans et Olivier Baudon
Labri - Université Bordeaux 1
Mondeca - Paris

10 janvier 2002

INTRODUCTION

Motivations

Contrat Mondeca-Labri :

Possibilité d'intégrer des algorithmes de théorie des graphes comme outils d'analyse de données dans un logiciel de structuration, publication, représentation et analyse d'informations de gestion.

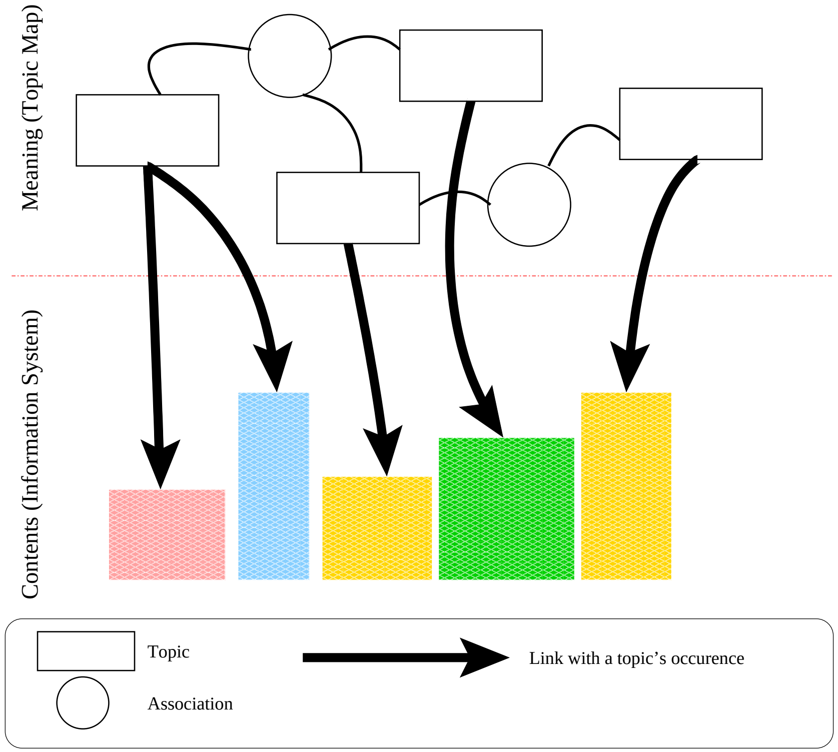
et en particulier

Proposition d'algorithmes pour réaliser une partition et trouver une entité remarquable, à l'intérieur d'un réseau d'entité et de relations de gestion sans domaine d'application spécifique.

Norme XML - Topic Map

En 1990, Mondeca participe à la normalisation du format XML-Topic Map et l'adopte pour son logiciel.

Réseau sémantique



XTM DTD pour Palm Navigator

```
<!-- XML Topic Map DTD  for Palm Navigator -->
<!-- XML Topic Map (XTM) DTD for Palm Navigator, Version 1.0

     This DTD is a subset of the XTM DTD version 1.0
     XML Topic Map (XTM)
       Copyright 2000 TopicMaps.Org, All Rights Reserved.

        Authors:    Group Mondeca
                    (DESS GL Université Bordeaux I)
        Version:    v1.0
-->
<!-- Use this URI to identify the default XTM namespace:

         "http://www.topicMaps.org/xtm/1.0/"

     Used to identify the XLink namespace:
```

```
         "http://www.w3.org/1999/xlink"
-->

<!-- topicMap: Topic Map document element -->

<!ELEMENT topicMap
   ( topic | association )*
>
<!ATTLIST topicMap
   id ID #IMPLIED
   xmlns CDATA #FIXED 'http://www.topicmaps.org/xtm/1.0/'
   xmlns:xlink CDATA #FIXED 'http://www.w3.org/1999/xlink'
   xml:base CDATA #IMPLIED
>

<!-- topic: Topic element -->
```

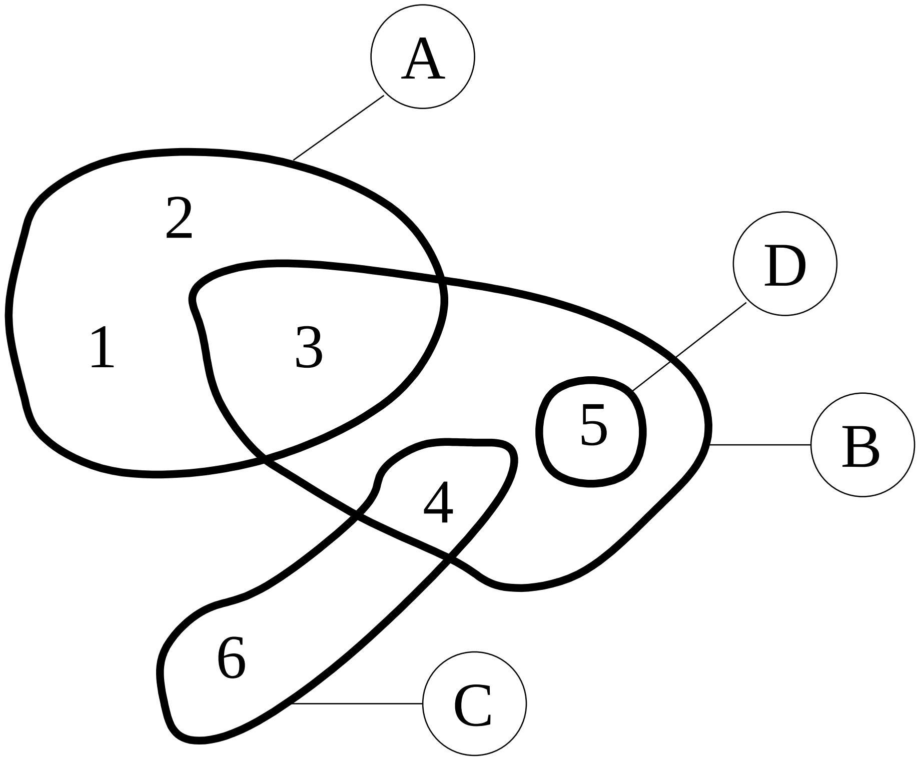
```
<!ELEMENT topic
   ( instanceOf*, ( baseName | occurrence )* )
>
<!ATTLIST topic
   id                ID        #REQUIRED
>

<!-- instanceOf: Points To a Topic Representing a Class -->

<!ELEMENT instanceOf  topicRef >
<!ATTLIST instanceOf EMPTY >

<!-- topicRef: Reference to a Topic Element -->

<!ELEMENT topicRef  EMPTY >
<!ATTLIST topicRef
```

```
   xlink:href      CDATA     #REQUIRED
>

<!-- baseName: Topic or Occurrence Base Name -->

<!ELEMENT baseName  baseNameString >
<!ATTLIST baseName EMPTY >

<!-- baseNameString: Base Name String Container -->

<!ELEMENT baseNameString  ( #PCDATA ) >
<!ATTLIST baseNameString EMPTY >

<!-- occurrence: Resources Regarded as an Occurrence -->

<!ELEMENT occurrence
   ( instanceOf?, baseName?, resourceData )
```

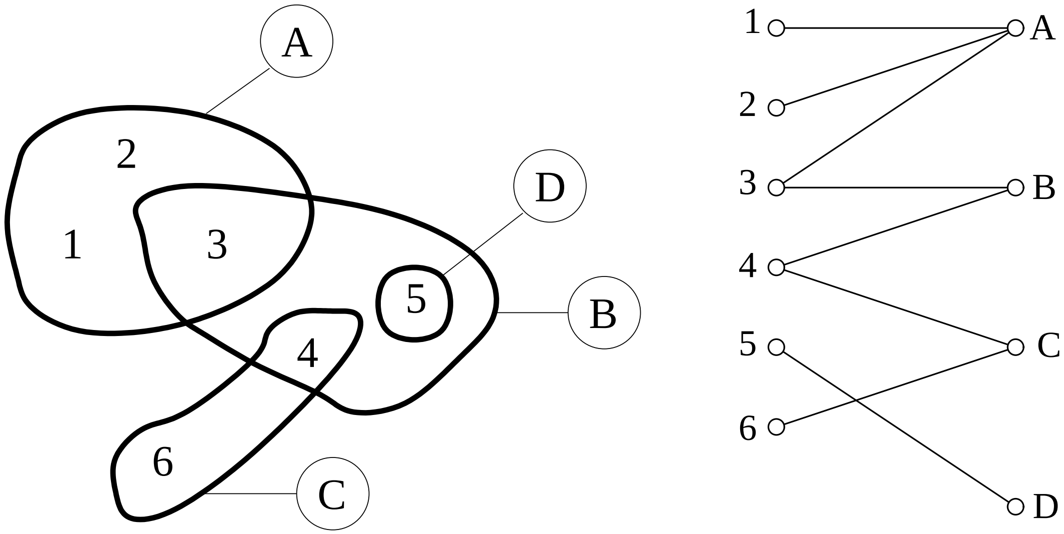
```
>
<!ATTLIST occurrence EMPTY >

<!-- resourceData: Container for Resource String -->

<!ELEMENT resourceData  ( #PCDATA ) >
<!ATTLIST resourceData EMPTY >

<!-- association: Topic Association  -->

<!ELEMENT association
   ( instanceOf?, member+ )
>
<!ATTLIST association
   id              ID        #REQUIRED
>
```

```
<!-- member: Member in Topic Association -->

<!ELEMENT member
   ( roleSpec?, topicRef+ )
>
<!ATTLIST member EMPTY >

<!-- roleSpec: Points to a Topic Serving as an
Association Role -->

<!ELEMENT roleSpec topicRef >
<!ATTLIST roleSpec EMPTY >

<!-- end of DTD XML Topic Map (XTM) for Palm
Navigator -->
```

Hypergraphe



Graphe représentatif



Topic Map *"Musique classique"*

219 116 topics et associations

1 496 592 rôles

Pascal Auillans et Olivier Baudon
Labri - Université Bordeaux 1
Mondeca - Paris

FRACTIONNEMENT

Définitions

On appelle **fractionnement (*clustering*) d'un graphe** $G = (V, E)$ un ensemble $\mathcal{S} = \{S_1, \ldots, S_k\}$ de parties de V dont l'union est égale à V.

Si les éléments de $\mathcal{S}$ sont tous disjoints deux à deux, ils forment une partition de V. On parlera alors de **partitionnement d'un graphe**.

Selon le cas, le sous-graphe induit par un élément S_i de $\mathcal{S}$ sera appelé une **fraction** ou une **partie** de G.

Domaines d'études

- Classification
- Equilibrage de charges en calcul distribué
- Réseaux d'interconnexion
- Etude théorique de la k-domination
- Visualisation de graphes
- Analyse factorielle
- Approches aléatoires

Approches

Fractionnement en k sous-ensembles, en optimisant un critère :
$\rightarrow$ partition en k parties équilibrées avec un nombre minimum d'arêtes entre les parties

Partition en un nombre minimum de fractions, chaque partie devant respecter un critère donné :
$\rightarrow$ diamètre borné
$\rightarrow$ rayon borné

Problème

La plupart des algorithmes et heuristiques proposés sont au moins en $O(n^2)$, en utilisant en particulier la matrice de distances

Choix

Fractionnement en un nombre minimum de fractions de rayon borné.
$\rightarrow$ élection naturelle d'un sommet.

Théorème : le problème de la k-domination est NP-complet, y compris en se restreignant aux graphes bipartis

Algorithmes

1. *inspiré d'une heuristique de Johnson (1974)*. A chaque itération, on construit une boule de rayon k à partir du sommet de poids le plus élevé. Le poids d'un sommet dépend de sa distance par rapport aux centres déja choisis.

2. *inspiré d'une heuristique de Eilam, Gavoille et Peleg (1998)*. On construit des boules à partir de centres choisis parmis les sommets les plus "couverts" à l'étape précédentes. Les centres de la première étape sont choisis aléatoirement.

3. *algorithme de Henning 1998, prouvant un minorant pour le nombre de boules maximum.* On choisit un arbre couvrant. A chaque étape, on calcule une chaîne de longueur égale au diamètre de l'arbre et on choisit comme centre le sommet à distance k d'une extrémité de cette chaîne. On réitère en supprimant de l'arbre le centre choisi ainsi que les sommets rattachés à l'arbre par un sommet de la chaîne comprise entre le sommet choisi et l'extrémité correspondante.

Graphes de tests

1. Grille 1000x300 ($n = 300000, m = 600000$)
2. Graphe aléatoire à 300 000 sommets et 600 000 arêtes
3. Graphe issu de la topic map “musique classique” ($n = 219116, m = 1496592$)

Grille 1000 x 300

n = 300 000, m = 600 000, boules de rayon 50

Algorithme	Temps	Nb. boules	Min	Max	Moy
1	3	74	1	2	1
2	42	240	1	3	3
3	697	5855	27	74	60

Graphe aléatoire

n = 300 000, m = 600 000, boules de rayon 5

Algorithme	Temps	Nb. boules	Min	Max	Moy
1	1425	11083	1	43	10
2	822	12157	1	37	9
3	1826	16949	1	1314	74

Graphe issu d'une topic map

n = 219 116, m = 1 496 592, boules de rayon 5

Algorithme	Temps	Nb. boules	Min	Max	Moy
1	58	373	1	57	3
2	76	408	1	130	8
3	1473	1105	436	919	495

Problèmes rencontrés

Qualité de la topic map

Chaînes pendantes

Choix de la fonction de poids et de l'initialisation

Pertinence de la notion de distance dans un graphe

Utilisation de la récursivité (rayon limite)

TRAVAUX EN COURS

Graphes t-couvrants

Un graphe partiel H de G est dit ***t*-couvrant** si et seulement si $\forall x, y \in V, d_H(x, y) \leq t \times d_G(x, y)$.

Deux approches :
soit on se donne t et on cherche le graphe partiel t-couvrant ayant un nombre minimum d'arêtes,
soit on se donne une propriété sur H et on cherche celui qui donne la plus petite valeur de t.

Soit t le plus petit réel tel que H est un graphe t-couvrant de G. t est appelé le **facteur d'étirement** de H par rapport à G.

Théorèmes (Cai)

Un arbre couvrant T de G est t-couvrant ssi
$\forall (x, y) \in E(G) \backslash E(T), d_T(x, y) \leq t \times d_G(x, y)$

L'existence d'un arbre t-couvrant est un problème NP-complet pour $t \geq 4$. C'est linéaire pour $t = 2$ et la complexité n'est pas connue pour $t = 3$.

Calculer le facteur d'étirement d'un arbre est linéaire.

Approche algorithmique

Partir d'un arbre de plus courtes chaînes avec un bon facteur d'étirement et rajouter les arêtes qui préserve la planarité du dessin d'arbre et qui diminuent le plus le facteur d'étirement.