Graphes médians, FCA et phylogénie

Master 2 Sciences Cognitives et Médias Numériqes

Baptiste Mounier

UFR Mathématiques et Informatique

LORIA

09 avril 2018 - 08 août 2018

Sommaire

Sommaire

1. Contexte
   - LORIA
   - ORPAILLEUR

LORIA

- **L**aboratoire **L**orrain de **R**echerche en **I**nforatique et ses **A**pplications
- UMR 7503 : CNRS, Université de Lorraine et Inria
- Création en 1997
- Fédération Charles Hermite
- 5 départements et 28 équipes - 400 personnes

ORPAILLEUR

- Équipe ORPAILLEUR
- 4ème département : Traitement automatique des langues et des connaissances
- Miguel Couceiro, Alian Gély, Amedeo Napoli
- Exploration de connaissances dans des bases de données

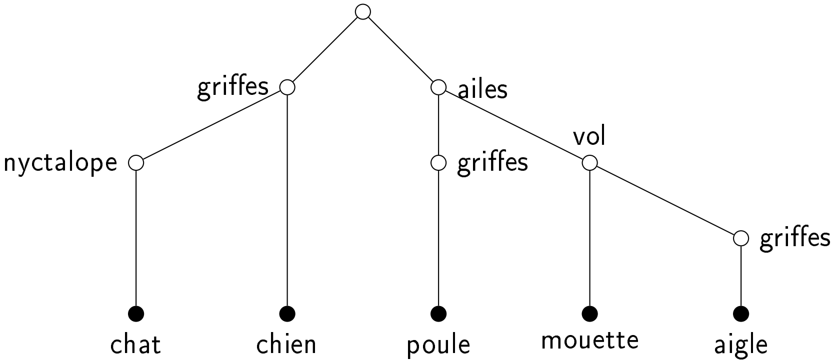
Sommaire

2 Motivations
- Phylogénie
- Graphes médians
- Analyse de Concepts Formels

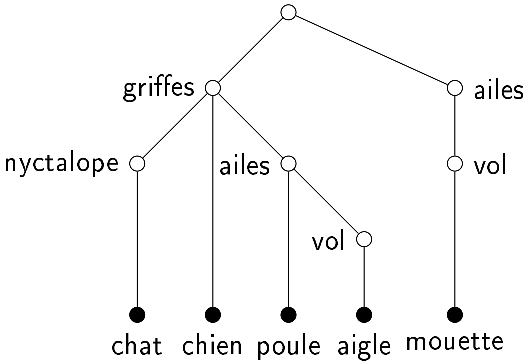
Données : matrice binaires espèce/caractéristique

	griffes	*ailes*	*nyctalope*	*vol*
chat	x		x	
chien	x			
aigle	x	x		x
mouette		x		x
poule	x	x		

Représentation : arbres

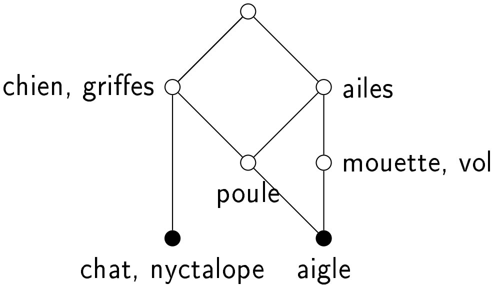


Représentation : arbres



Un outil : Graphes médians

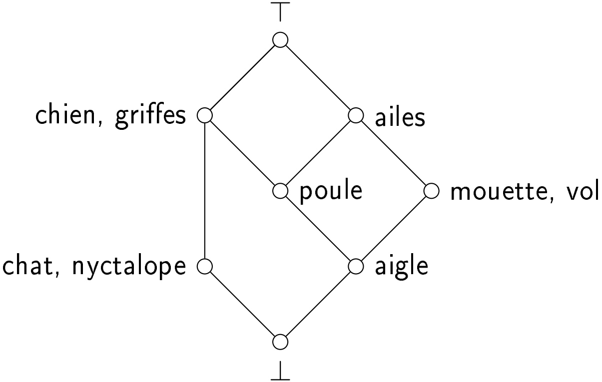
- Hans-Jürgen Bandelt
- Graphe médian
- Intégralité des arbres parcimonieux



Une approche : Analyse de Concepts Formels

- Les matrices binaires sont centrales
- Catégorisation d'individus suivant leurs caractéristiques

Une approche : Analyse de Concepts Formels



Une approche : Analyse de Concepts Formels

- Uta Priss
- Outils et la communauté de la FCA pour la phylogénie
- Liens entre treillis distributifs et graphes médians
- Proposition de méthode : transformer un treillis quelconque en un graphe médian

Une approche : Analyse de Concepts Formels

- Orpailleur (Yacine Namir)
- Implémentation d'une solution
- Problèmes sur certains types de cas

Sommaire

3. Notions nécessaires
   - Analyse de Concepts Formels
   - Ensemble ordonné
   - Treillis
   - Graphe médian

Contexte $C(O, A, I)$

$O = \{chat, chien, aigle, mouette, poule\}$
$A = \{griffes, ailes, nyctalope, vol\}$
$I = \{(chat, griffes), (chat, nyctalope), (chien, griffes), \ldots\}$

	griffes	*ailes*	*nyctalope*	*vol*
chat	x		x	
chien	x			
aigle	x	x		x
mouette		x		x
poule	x	x		

Connexion de Gallois

Definition

Pour $X \subseteq O$ et $Y \subseteq A$, on définie :

- $X' = \{y \in A : (x, y) \in I, \forall x \in X\}$
- $Y' = \{x \in O : (x, y) \in I, \forall y \in Y\}$

Connexion de Gallois

	griffes	*ailes*	*nyctalope*	*vol*
chat	x		x	
chien	x			
aigle	x	x		x
mouette		x		x
poule	x	x		

$\{chat, aigle\}' = \{griffes\}$
$\{griffes, ailes\}' = \{aigle, poule\}$

Contexte clarifié

- À partir d'un contexte
- Aucune redondance
- Conservation de la structure
- Perte uniquement de labels

Definition

Un contexte $C(O, A, I)$ est clarifié si et seulement si :

- $\forall x1, x2 \in O$ si $x1' = x2'$ alors $x1 = x2$
- $\forall y1, y2 \in A$ si $y1' = y2'$ alors $y1 = y2$

Contexte clarifié

	a	b	c	d	e
1	x			x	
2	x				
3	x	x	x		x
4		x	x		x
5	x				
6	x		x		

	a	b	c	d
1	x			x
2	x			
3	x	x	x	
4		x	x	
6	x		x	

Contexte réduit

- À partir d'un contexte clarifié
- Aucune information recalculable
- Conservation de la structure
- Perte uniquement de labels

Definition

Un contexte clarifié $C(O, A, I)$ est réduit si et seulement si :

- $\forall x \in O, \forall X \subseteq O$, si $x' = X'$ alors $x \in X$
- $\forall y \in A, \forall Y \subseteq A$, si $y' = Y'$ alors $y \in Y$

Contexte réduit

	a	b	c	d
1	x			x
2	x			
3	x	x	x	
4		x	x	
6	x		x	

	a	b	c	d
1	x			x
3	x	x	x	
4		x	x	
6	x		x	

Concept

Definition

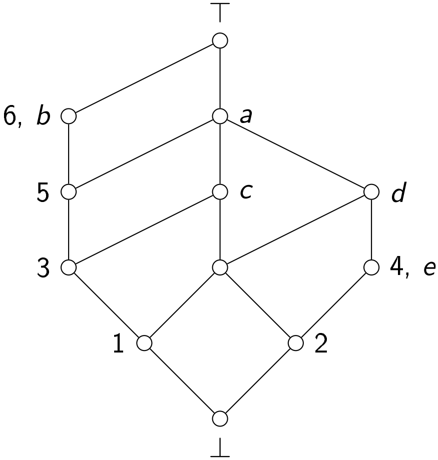
Soit un concept c, $X \subseteq O$ et $Y \subseteq A$:

- $c = \{X, Y\}$
- $\forall x \in O,\ x \in X \Leftrightarrow x' \subseteq Y$
- $\forall y \in A,\ y \in Y \Leftrightarrow y' \subseteq X$

	a	b	c	d
1	x			x
3	x	x	x	
4		x	x	
6	x		x	

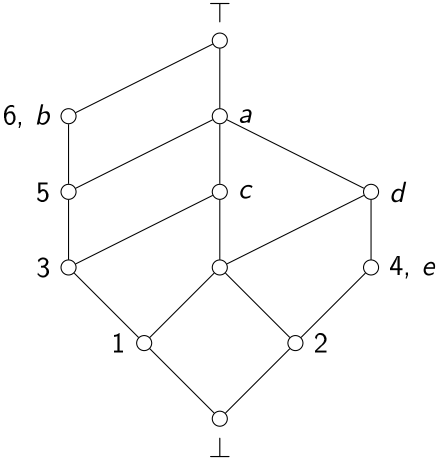
Treillis de concept

	a	b	c	d	e
1	x	x	x	x	
2	x		x	x	x
3	x	x	x		
4	x			x	x
5	x	x			
6		x			



Relations flèches

	a	b	c	d	e
1	x	x	x	x	$\updownarrow$
2	x	$\updownarrow$	x	x	x
3	x	x	x	$\updownarrow$	
4	x	$\uparrow$	$\updownarrow$	x	x
5	x	x	$\updownarrow$	$\uparrow$	
6	$\updownarrow$	x			

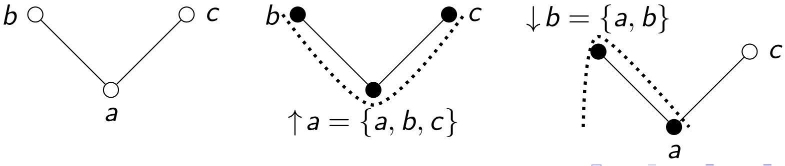


Ensemble ordonné

Definition

Soit un ensemble ordonné $(P, \leq)$, $X \subseteq P$ et $Y \subseteq P$:

- $\forall x, y \in P$, on a $x \leq y$ ou $y \leq x$
- $\uparrow X = Y \Rightarrow x \leq y \ \forall x \in X, \forall y \in Y$
- $\downarrow X = Y \Rightarrow y \leq x \ \forall x \in X, \forall y \in Y$

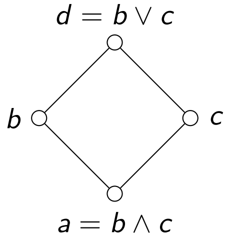


Treillis

Definition

Soit un treillis $(T, \leq, \vee, \wedge)$:

- $\forall x, y \in T, \exists z \in T : x \vee y = z$
- $\forall x, y \in T, \exists z \in T : x \wedge y = z$



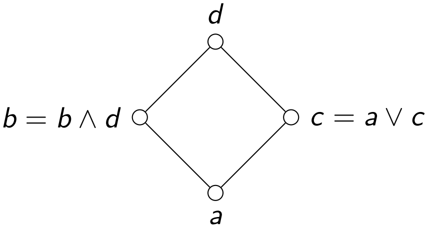
- Ensemble ordonné
- Opération de borne supérieur $\vee$
- Opération de borne inférieur $\wedge$

Irréductibles

Definition

Soit un treillis $(T, \leq, \vee, \wedge)$, $x, y, z \in T$:

- x est $\vee$-irreductible ssi $x \vee y = z \Rightarrow x = z$
- x est $\wedge$-irreductible ssi $x \wedge y = z \Rightarrow x = z$



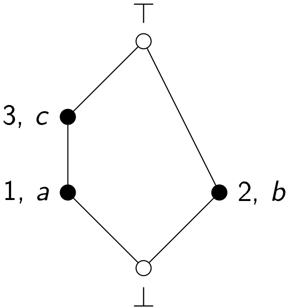
Treillis distributif

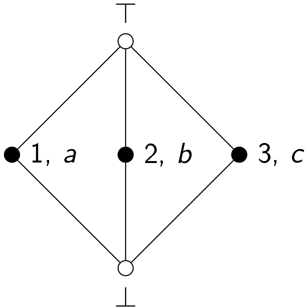
Definition

Un treillis $(T, \leq, \vee, \wedge)$ est distributif si et seulement si au moins l'une des trois conditions équivalentes suivantes est vérifiée :

- $(x \wedge y) \vee (x \wedge z) \vee (y \wedge z) = (x \vee y) \wedge (x \vee z) \wedge (y \vee z)$
- Le treillis ne possède ni N_5 ni M_3 en guise de sous treillis
- Le contexte réduit contient une seule relation flèche double par ligne et par colonne

Treillis distributif

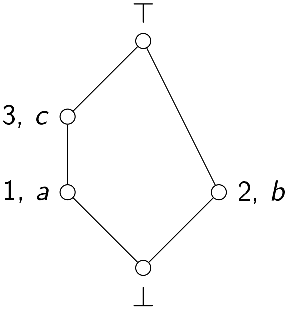


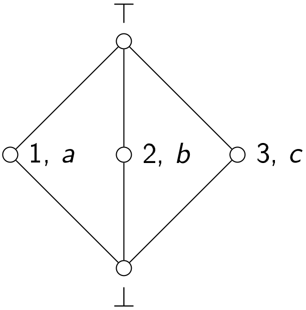


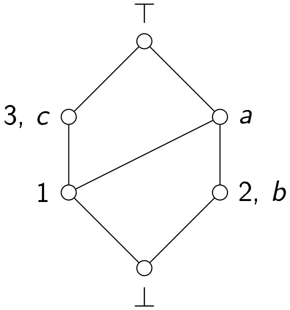
Graphe médian

- Un unique nœud commun à tous les chemins les plus courts entre les éléments d'un triplets de nœuds
- Liens forts avec les treillis distributifs

Graphe médian







Sommaire

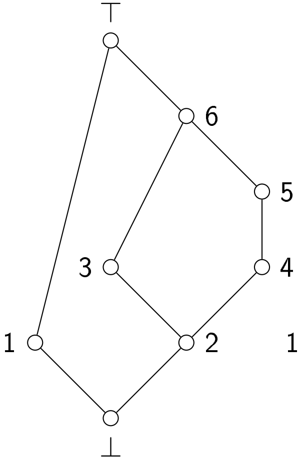
4 Existant
- Méthode
- Algorithme
- Problèmes

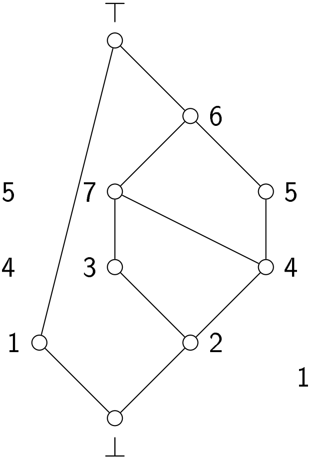
Méthode

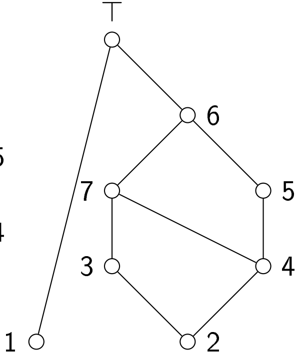
- Treillis distributif $\Rightarrow$ graphe médian
- Plus proche possible du treillis d'origine

- Rendre distributifs les treillis des filtres des atomes

Méthode

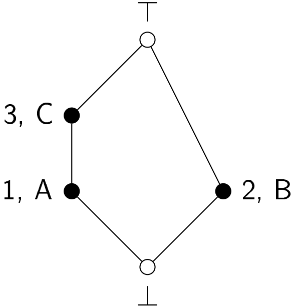






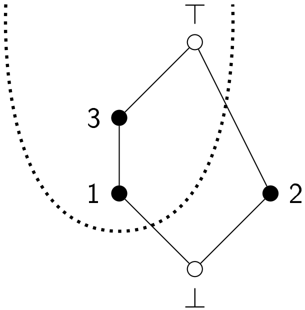
Algorithme

	A	B	C
1	x		x
2		x	
3			x



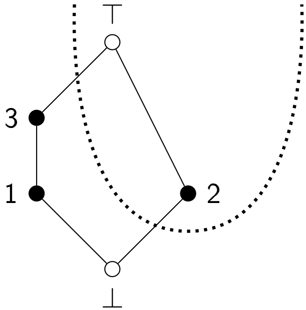
Algorithme

	m1
1	
2	x
3	



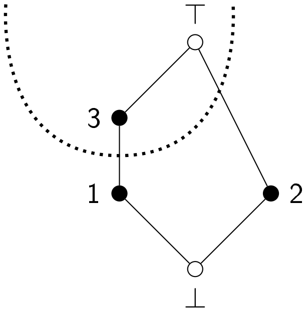
Algorithme

	m1	m2
1		x
2	x	
3		x



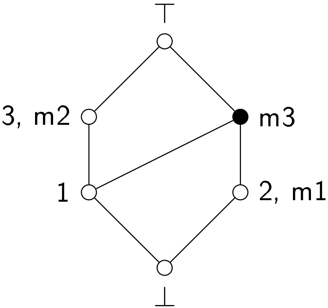
Algorithme

	m1	m2	m3
1		x	x
2	x		x
3		x	

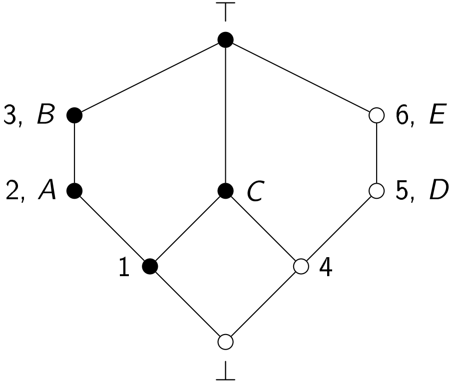


Algorithme

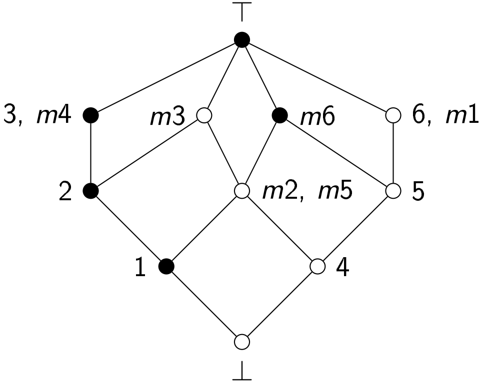
	m1	m2	m3
1		x	x
2	x		x
3		x	



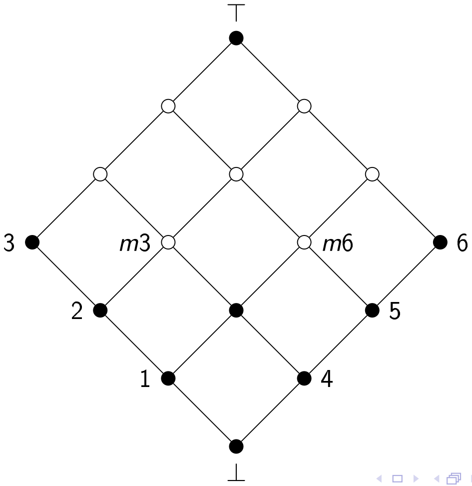
Problèmes



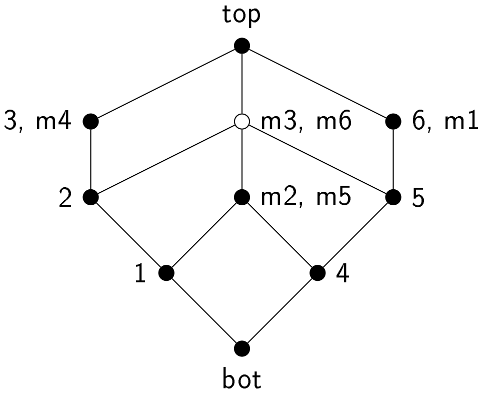
Problèmes



Problèmes

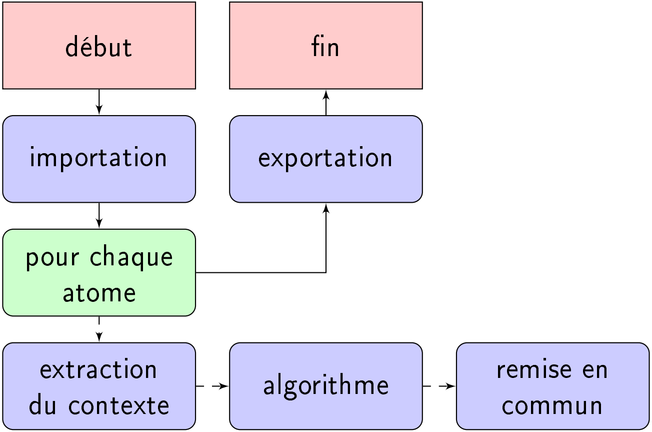


Problèmes



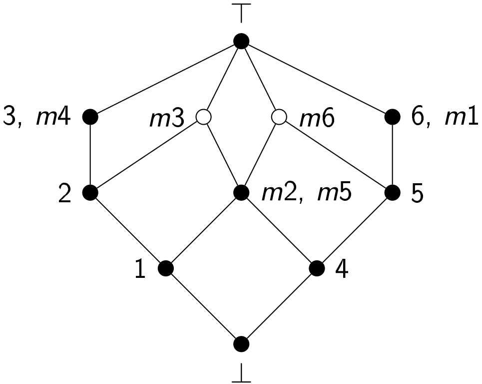
Sommaire

Implémentation : workflow



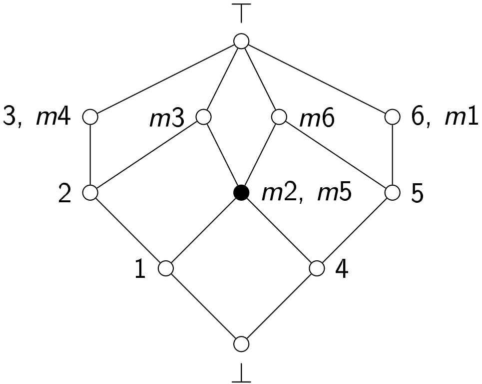
Fusion

Les deux nœuds à fusionner ne doivent pas faire partis des nœuds du treillis de départ.



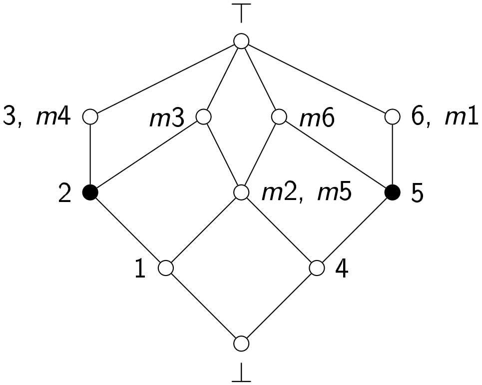
Fusion

Les deux nœuds à fusionner doivent être en couverture d'un nœud présent dans au moins deux filtres d'atomes.

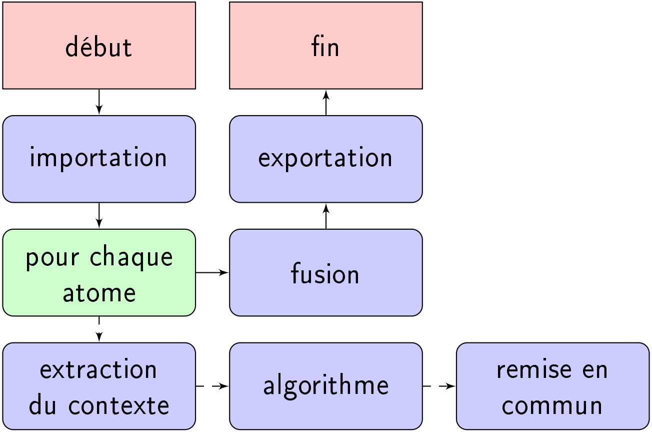


Fusion

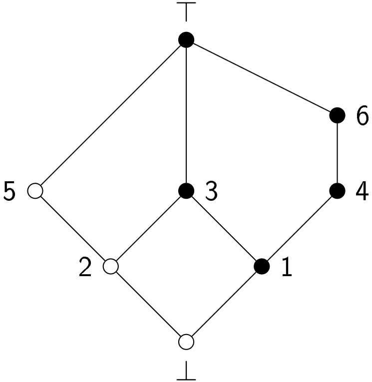
Les deux nœuds à fusionner ne doivent pas avoir de nœuds en commun dans leurs idéaux une fois que les idéaux du nœud en commun (celui dont il est question dans la condition précédente) leur sont retirés.



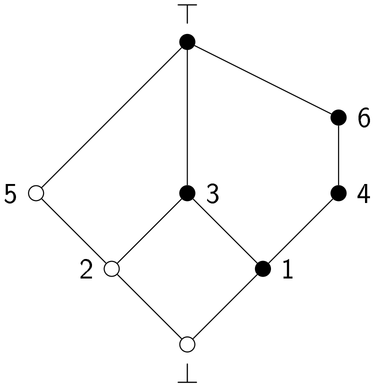
Fusion : workflow

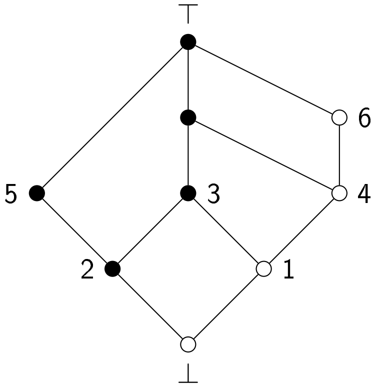


Système de boucle

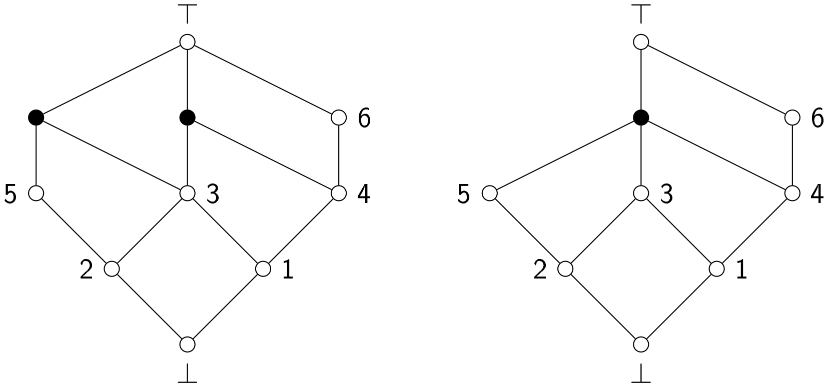


Système de boucle

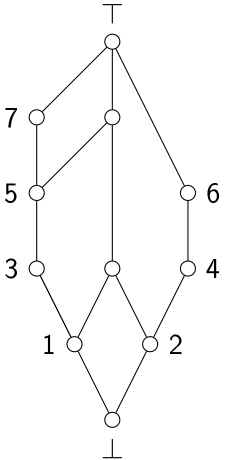


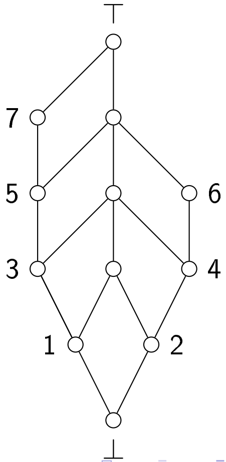


Système de boucle

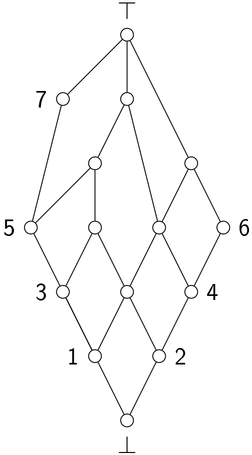


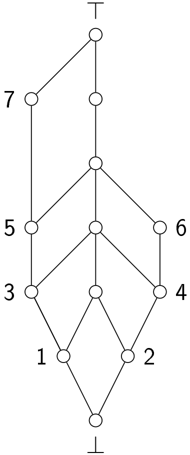
Apparition de chaines



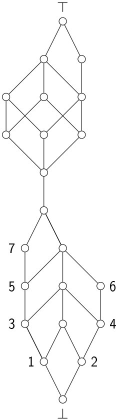


Apparition de chaines





Apparition de chaines



Système de boucle

- Fusion au fur et à mesure
- Fusion après tout le système de boucle

Apparition de chaines

- Solution bancale
- Complication de la justification par preuves
- Recherche d'une solution plus adaptée

Optimalité du treillis

- Définition d'un système métrique de comparaison
- $D_{T1-T2}(a, b)$: a est le nombre de nœuds ajoutés et b le nombre d'arêtes ajoutées
- Quantifier la performance de la méthode

Optimisation du programme

- Gain de performance logicielle
- Simplification de certains morceaux de code
- Réduction des possibilités de bogues

Sommaire

6 Bilan
- Contribution
- Stage

Contribution

- Programme générant un treillis prêt à être converti en graphe médian pour la plupart des cas
- Proposition de trois conditions pour effectuer la fusion de nœuds
- Nécessitant d'y apporter des preuves

Stage

- Gains techniques classiques (Python, Latex, etc)
- Goût pour l'échange et le débat autour de notions
- Goût pour cette recherche constante d'amélioration même si on en n'est pas le bénéficiaire direct

Dépôt github du stage.
Disponible sur : `https://github.com/BaptisteMounier/FCA_medianGraph/`.
Consulté de avril à août 2018.

Hans-Jürgen Bandelt, Peter Forster, and Röhl Arne.
Median-joining networks for inferring intraspecific phylogenies.
Molecular Biology and Evolution, page 12, 1998.

Hans-Jürgen Bandelt and Jarmila Hedlíková.
Median algebras.
Discrete Mathematics, page 30, 1980.

Sebastian Bank.
Documentation graphviz pour python.
Disponible sur : `https://pypi.org/project/graphviz/`.
Consulté de avril à août 2018.

Merci de votre attention.
Avez-vous des question ?