Analyse de Concepts Formels, distributivité et modèles de graphes médians pour la phylogénie

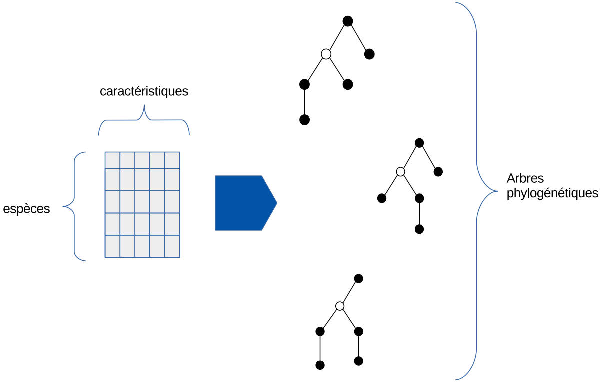
A. Gély M. Couceiro A. Napoli

LORIA (INRIA - CNRS - Université de Lorraine)

Conférence SFC 2019, Nancy, 5 Septembre 2019

Plan

Phylogénie - Liens inter-espèces



On recherche des arbres parcimonieux, Plusieurs arbres peuvent être solution

Arbtres phylogénétiques

- Mettent en avant les filiations inter-espèces
- Permettent de suivre l'évolutions des mutations
- Ne sont pas uniques

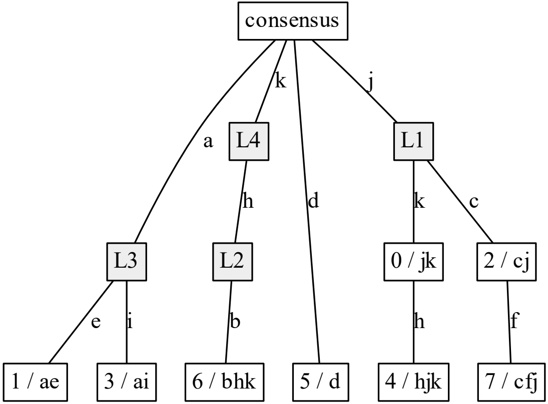
Données phylogénétiques (1)

	a	b	c	d	e	f	g	h	i	j
0							×			×
1	×				×					
2			×							×
3	×								×	
4							×	×		×
5				×						
6		×						×		
7			×			×				×

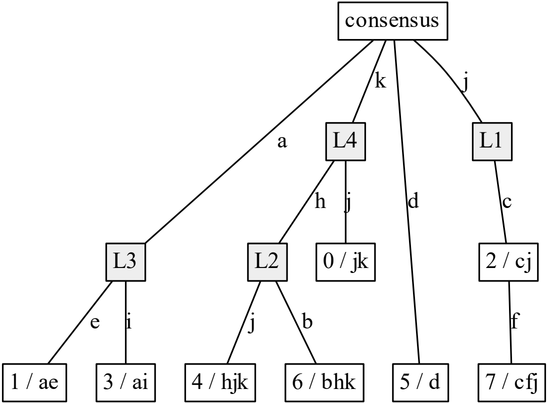
Etude de variabilité génétique d'une population de chasseurs-cueilleurs[1]

[1] Hans-Jürgen Bandelt, Vincent Macaulay, and Martin Richards (2000). "Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA". In: *Molecular phylogenetics and evolution* 16.1, pp. 8–28.

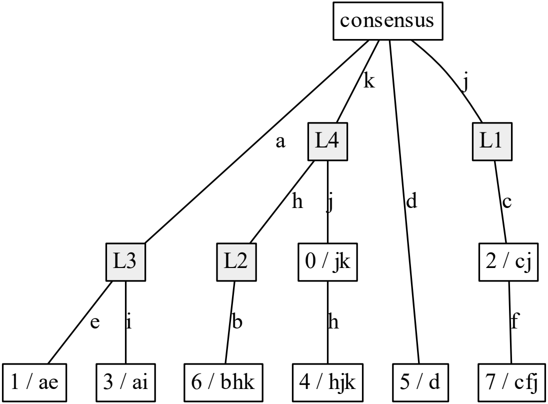
Arbres parcimonieux et graphes médians



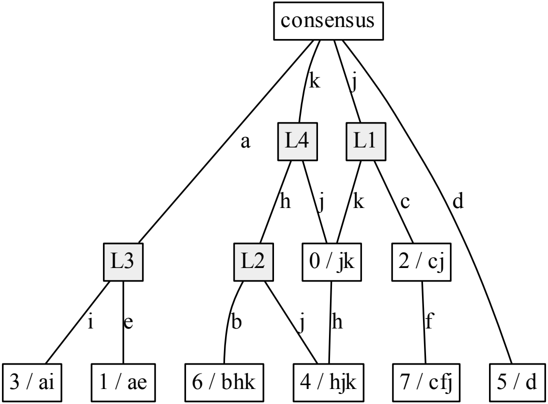
Arbres parcimonieux et graphes médians



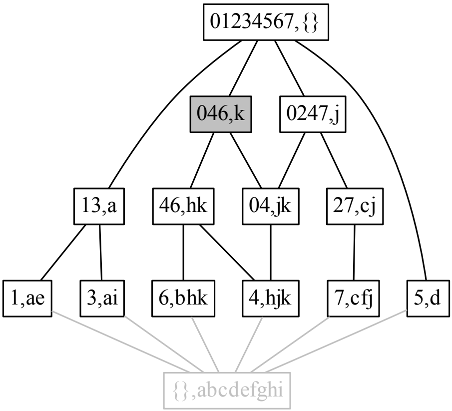
Arbres parcimonieux et graphes médians



Arbres parcimonieux et graphes médians

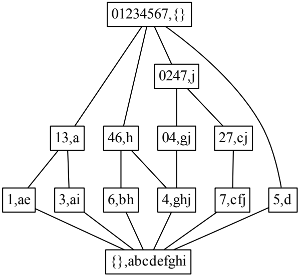


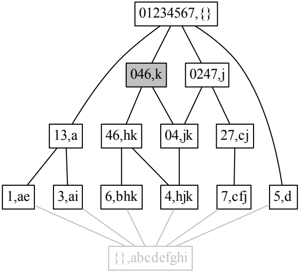
Arbres parcimonieux et graphes médians



Données phylogénétiques (2)

	a	b	c	d	e	f	g	h	i	j	k
0							×			×	×
1	×				×						
2			×							×	
3	×								×		
4							×	×		×	×
5				×							
6		×						×			×
7			×			×				×	


01234567,{}
0247,j
13,a
46,h
04,gj
27,cj
1,ae
3,ai
6,bh
4,ghj
7,cfj
5,d
{},abcdefghi


01234567,{}
046,k
0247,j
13,a
46,hk
04,jk
27,cj
1,ae
3,ai
6,bhk
4,hjk
7,cfj
5,d
{},abcdefghi

Graphe médian et Treillis Distributifs

- Liens forts entre *graphe médian* et *treillis distributifs*
- Treillis utilisés dans le cadre de la FCA
- U. Priss propose un algorithme pour utiliser les outils FCA en phylogénie.

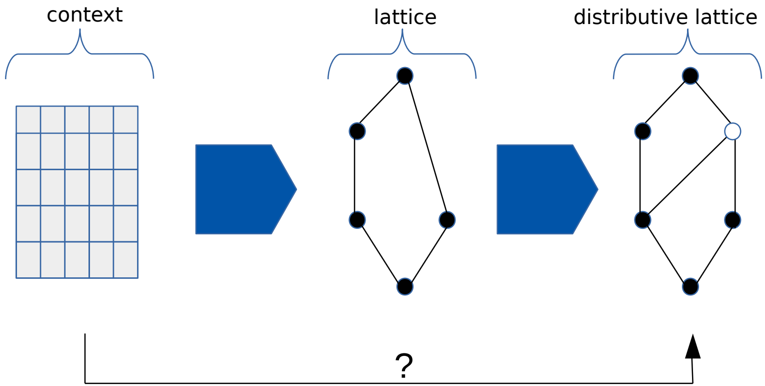
Remarque

Priss, 2013

> *"an algorithm for converting a concept lattice [into a median graph] consists of omitting the bottom node and then checking every principal filter for distributivity and turning it into a distributive lattice if it is not already one."*

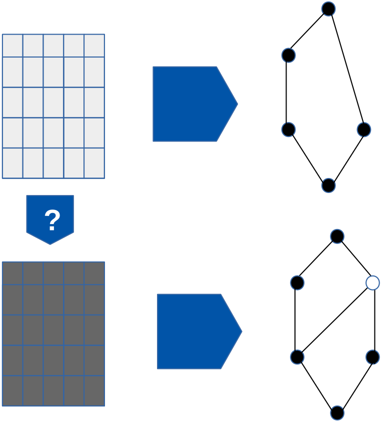
But

Construire le contexte d'un semi-treillis $\vee-$distributif sans construire le treillis

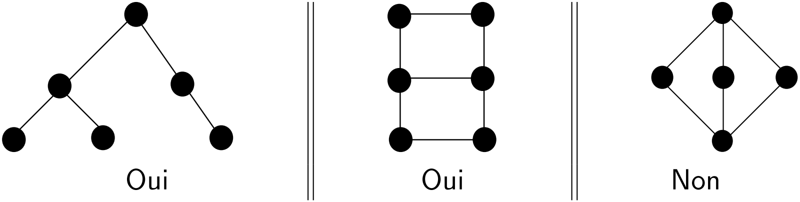


But

Construire le contexte d'un semi-treillis $\vee$−distributif sans construire le treillis



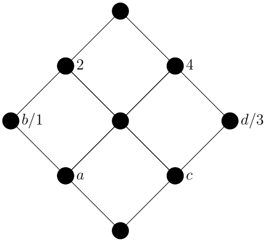
Graphe Médian
Définition



Definition

Graphe Médian $G = (V, E)$ tel que $\forall x, y, z \in V$, il existe un unique sommet qui est l'intersection des plus courts chemins entre (x, y), (y, z) et (x, z).

Treillis et Contexte



	1	2	3	4
a	×	×		×
b	×	×		
c		×	×	×
d			×	×

Definition

- $(T, \leq, \vee, \wedge)$ un treillis
- J l'ensemble des éléments $\vee$-irréductibles
- M l'ensemble des éléments $\wedge$-irréductibles
- On considère un contexte $(J, M, \leq)$

Definition

Un treillis est distributif ssi $\vee$ et $\wedge$ sont distributifs entre eux.

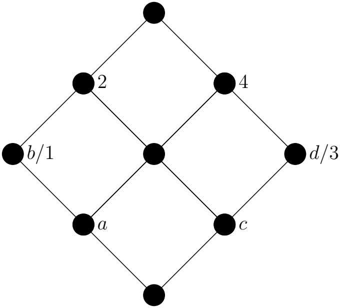
- $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$
- $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$

Théorème

Un treillis $(T, \leq, \vee, \wedge)$ est distributif ssi une (toutes) les définitions équivalentes sont satisfaites :

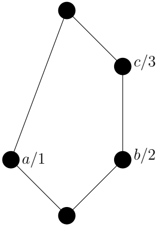
1. $(x \wedge y) \vee (y \wedge z) \vee (z \wedge x) = (x \vee y) \wedge (y \vee z) \wedge (z \vee x)$
2. *T contient exactement une double-flèche $\swarrow\!\!\nearrow$ par ligne et par colonne, sans autre flèches.*
3. *T ne contient ni N_5 ni M_3 comme sous-treillis*

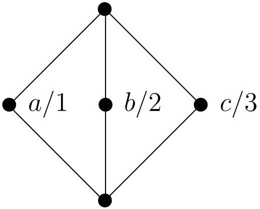
Treillis et relations flèches - Treillis distributifs.



	1	2	3	4
a	×	×	⤢	×
b	×	×		⤢
c	⤢	×	×	×
d		⤢	×	×

Treillis N_5 et M_3.

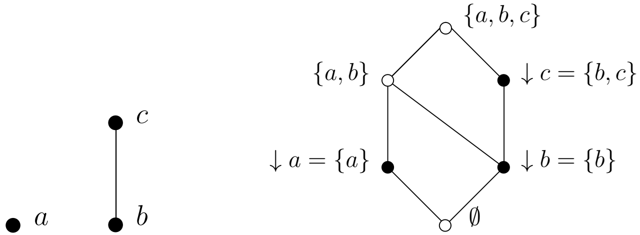




N_5	1	2	3
a	$\times$	$\searrow$	$\swarrow\!\!\!\!\nearrow$
b	$\swarrow\!\!\!\!\nearrow$	$\times$	$\times$
c	$\nearrow$	$\swarrow\!\!\!\!\nearrow$	$\times$

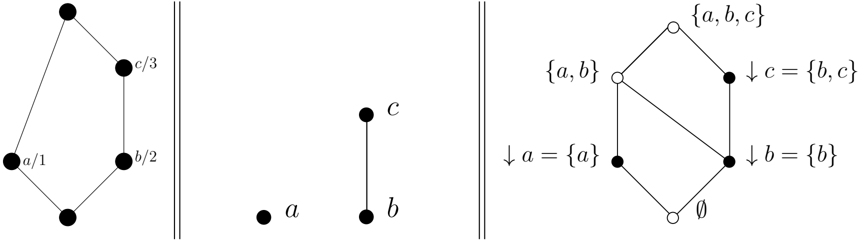
M_3	1	2	3
a	$\times$	$\swarrow\!\!\!\!\nearrow$	$\swarrow\!\!\!\!\nearrow$
b	$\swarrow\!\!\!\!\nearrow$	$\times$	$\swarrow\!\!\!\!\nearrow$
c	$\swarrow\!\!\!\!\nearrow$	$\swarrow\!\!\!\!\nearrow$	$\times$

- L'ensemble des idéaux d'ordre d'un ensemble ordonné est un treillis, appelé treillis des idéaux
- Les treillis des idéaux sont en bijection avec les treillis distributifs
- Plus précisément, un treillis distributif est en bijection avec le treillis des idéaux de $(J, \leq)$



Remarque

Pour un treillis non distributif T, il n'y a pas d'isomorphisme avec $\mathcal{O}(J(T))$.
Mais il existe un plongement de T vers $\mathcal{O}(J(T))$.



Idée

Utiliser un plongement pour obtenir un treillis distributif qui conserve les relations initiales entre les données.

Problème

Comment calculer le contexte de ce treillis distributif ?

Theorem (GW96)

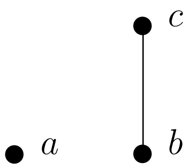
Soit $(P, \leq)$ un ensemble ordonné, Alors $C(P, P, \not\geq)$ est le contexte de $\mathcal{I}(P)$

Calcul du contexte du plongement d'un treillis dans un treillis distributif

Calcul de $(J, \leq)$

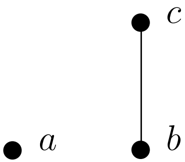
	1	2	3
a	×		
b		×	×
c			×

$$j_1 \leq j_2 \iff j_2' \subset j_1'$$



Calcul du contexte du plongement d'un treillis dans un treillis distributif

Calcul du contexte



	m_a	m_b	m_c
a		×	×
b	×		×
c	×		

Remarque

Priss, 2013

> *"an algorithm for converting a concept lattice [into a median graph] consists of omitting the bottom node and then checking every principal filter for distributivity and turning it into a distributive lattice if it is not already one."*

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

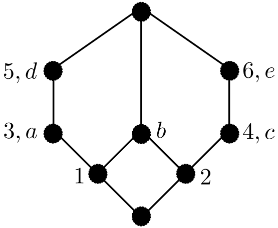
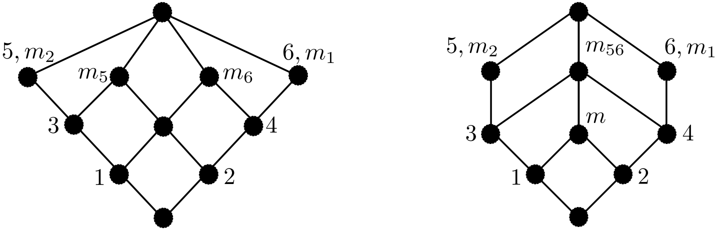


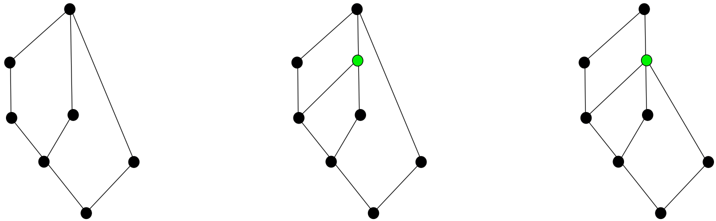
Figure: Contexte problématique et treillis des concepts associé

	m_2	m_5	m_1	m_6
2		×	×	×
3	×	×		
5	×			
1	×	×		×
4			×	×
6			×	



(a) (b)

Figure: (a) Solution obtenue après une approche locale et (b) solution optimale



(a) (b) (c)

- Un treillis (a) tel qu'il existe 2 $\vee$-semi-treillis (sans $\perp$) minimaux (b) and (c) dans lesquels plonger a

En résumé

- Algorithme qui produit un plongement dans un treillis distributif ;
- Retourne le contexte sans avoir besoin de construire le treillis ;
- Conserve un lien entre le treillis initial et le treillis distributif, via l'isomorphisme de $(J, \leq)$
- Recherche d'une solution optimale