Treillis en classification et fouille de données

Panorama de l'utilisation des treillis à Metz

Alain Gély
LITA, Université Paul Verlaine, Metz
gely@univ-metz.fr

Journée des Treillis Clermontois
Clermont-Ferrand
3 Décembre 2009

Organisation & thèmes de recherche

Université **P**aul **V**erlaine **M**etz

Laboratoire d'**I**nformatique **T**héorique et **A**ppliquée

Équipe Algorithmique et Optimisation

Thème "Fouille de données et extraction de connaissances"

Lydia Boudjeloud

- Visualisation de données
- Données évolutives
- Masses de données

François Brucker

http://francois.brucker.perso.centrale-marseille.fr/

- Classification
- Mathématiques discrètes
- Combinatoire

Alain Gély

http://www.lita.univ-metz.fr/~gely/

- Algorithmique
- Structures discrètes
- Treillis & représentations associées

collaborations sur un Contrat Plan Etat Région "Traitement et visualisation de grands ensembles de données"

Système de classes sur X

Famille $\mathcal{F}$ d'ensembles tq
$\{x\} \in F$ quel que soit $x \in X$
$X \in F$.

Système de fermeture sur X

Famille $\mathcal{F}$ d'ensembles tq
$Y,Z \in \mathcal{F}$ implique $Y \cap Z \in \mathcal{F}$
$X \in F$.

quitte à ajouter les intersections manquantes, on peut toujours ramener un système de classes à un système de fermeture.

Système de classes sur X

Famille $\mathcal{F}$ d'ensembles tq

$\{x\} \in F$ quel que soit $x \in X$

$X \in F$.

Système de fermeture sur X

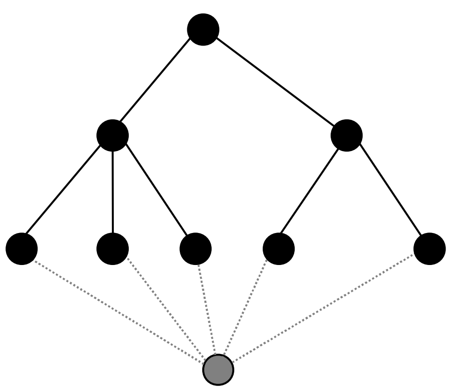
Famille $\mathcal{F}$ d'ensembles tq

$Y,Z \in \mathcal{F}$ implique $Y \cap Z \in \mathcal{F}$

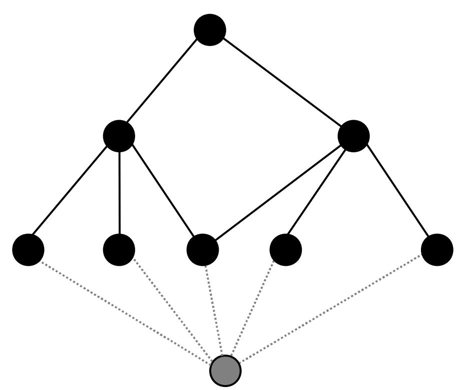
$X \in F$.

En fonction des propriétés que l'on cherche à obtenir sur le système de classe, on aura différents types de systèmes de fermeture.

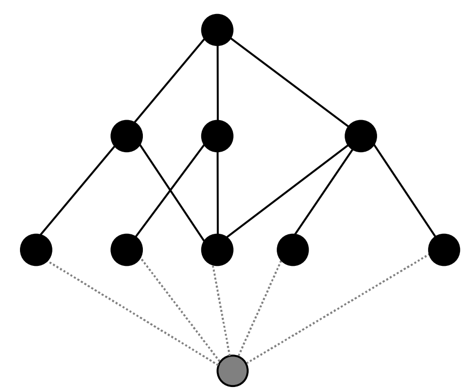
- Quasi Hiérarchies
- Pseudo Hiérarchies
- Hiérarchies
- ...



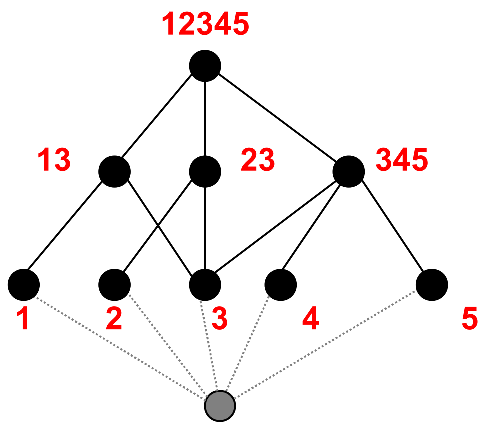
hiérarchie



Pseudo-hiérarchie



Quasi-hiérarchie



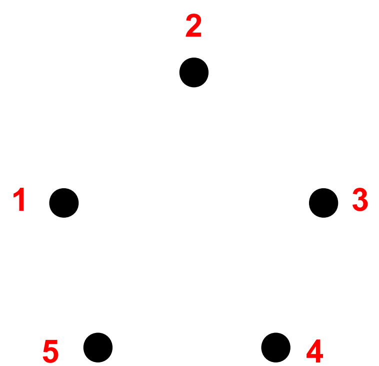
Quasi-hiérarchie

Les quasi-hiérarchies sont une frontière.

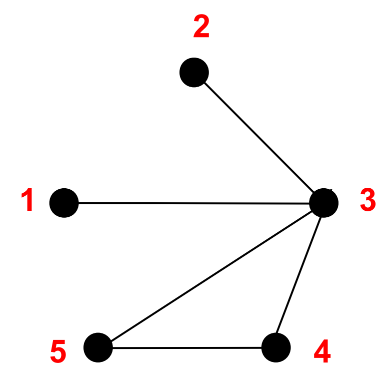
Correspondance entre **n'importe quel indice** et le système de classes, vu comme **famille de cliques maximales** de graphes seuils.

	1	2	3	4	5
1	0	2	1	2	2
2		0	1	2	2
3			0	1	1
4				0	1
5					0

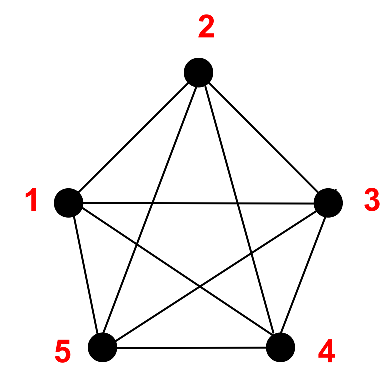
Matrice
de dissimilarité



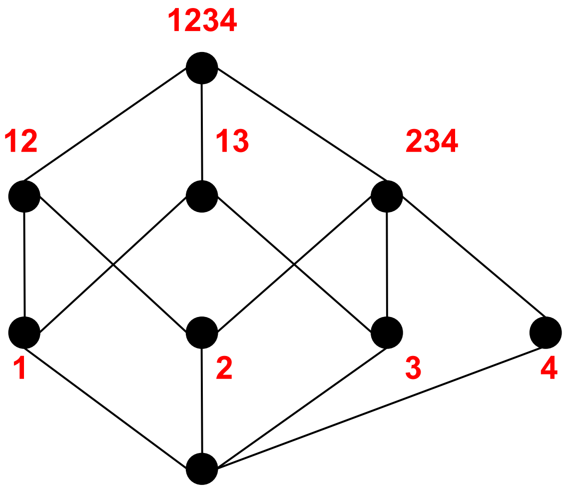
graphe seuil
≤ 0



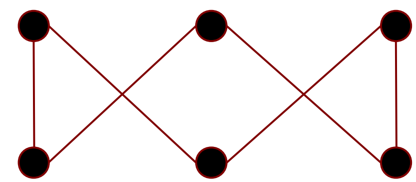
graphe seuil
≤ 1



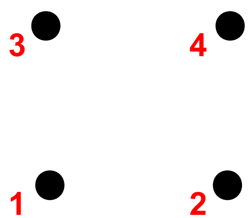
graphe seuil
≤ 2



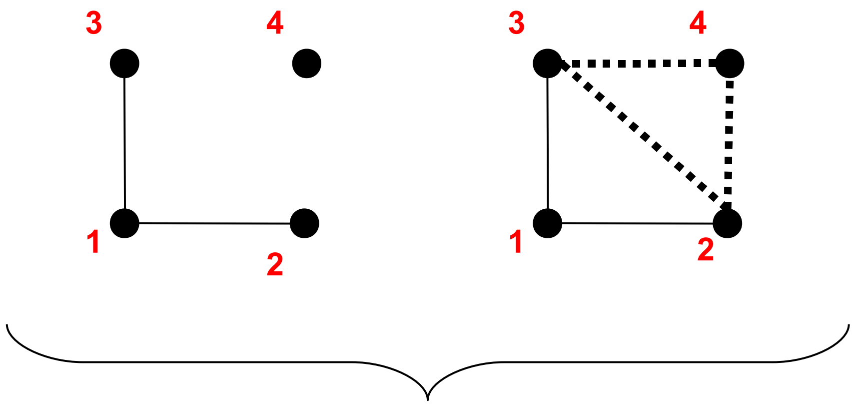
dans une **quasi-hiérarchie,** l'intersection de trois classes est égale à l'intersection de deux d'entre elles.



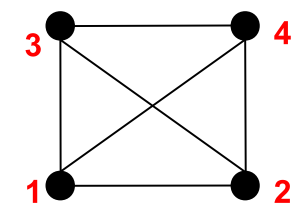
Couronne à 3 éléments
Configuration interdite



graphe seuil
≤ 0



impossibilité de retrouver les classes



graphe seuil
$\leq \infty$

quasi-hiérarchie

frontière indices / Cliques.
configuration interdite : 3-couronnes

système de classes parcimonieux

Introduit par F. Brucker comme généralisation des arbres de phylogénie

pseudo-hiérarchie

Diagramme de Hasse Planaire
(hypergraphes d'intervalles)

hiérarchie

Arbres (emboîtement des classes)

Étude des systèmes de classes parcimonieux

- En bijection avec la classe des **treillis démantelables** atomiques définie par Rival en 74
- Configuration interdite : n-couronne
- chaque classe peut être définie par 2 éléments
- Propriétés d'élimination

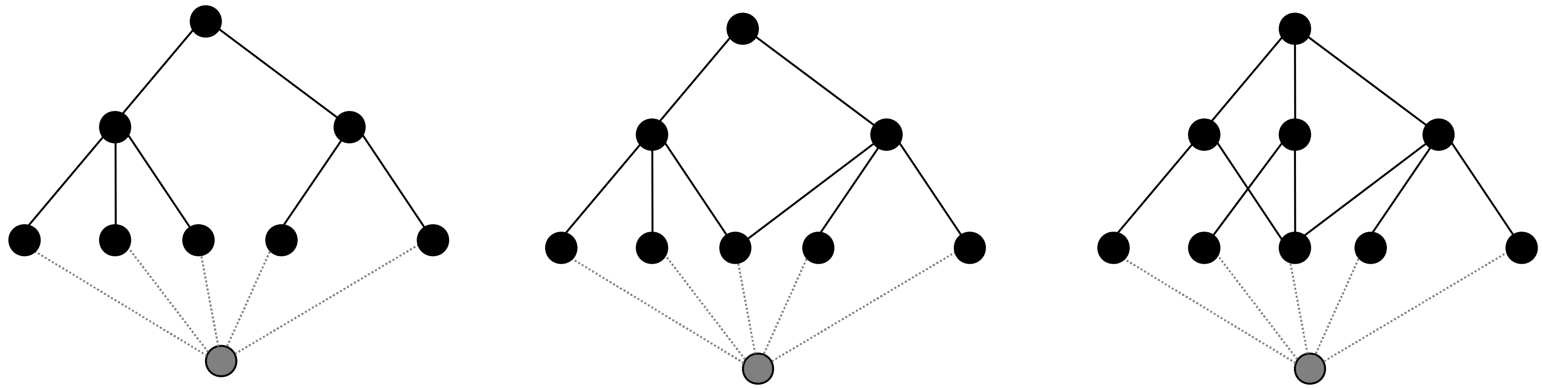
Aux précédentes journées treillis... (journées lorraines)

- Les systèmes de classes parcimonieux (alors appelés X-hyperarbres) sont les hyper-arbres stables par restrictions.
- Les hyper-arbres stables par restrictions sont en bijection avec les treillis démantelables.

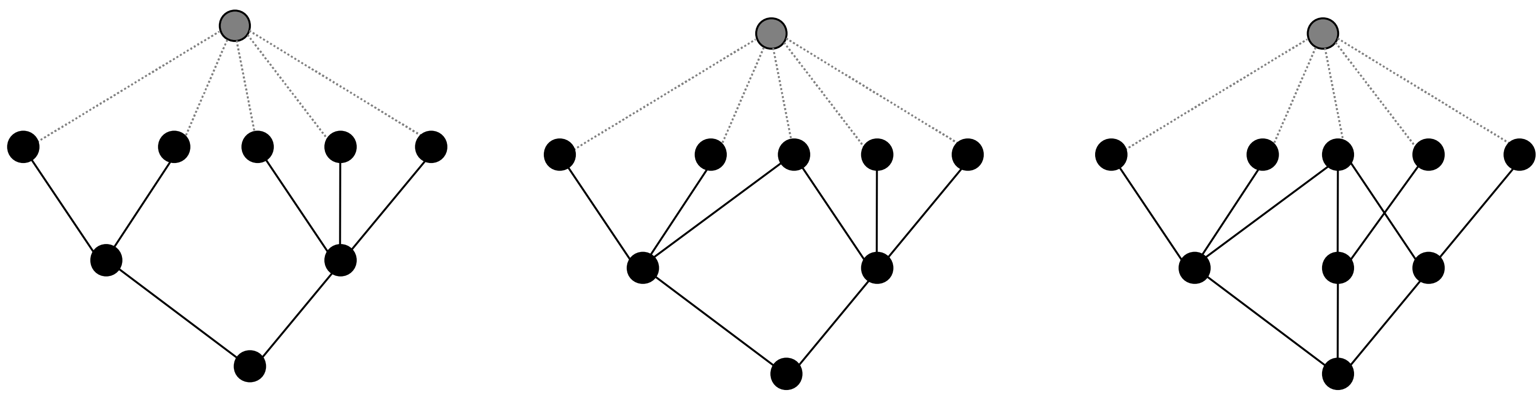
Conséquences

- Bonne propriétés pour la classification (on peut faire une classification identique sur une sous partie des espèces)
- Informations sur les interactions entre classes contenues dans l'ordre d'élimination (effeuillage)

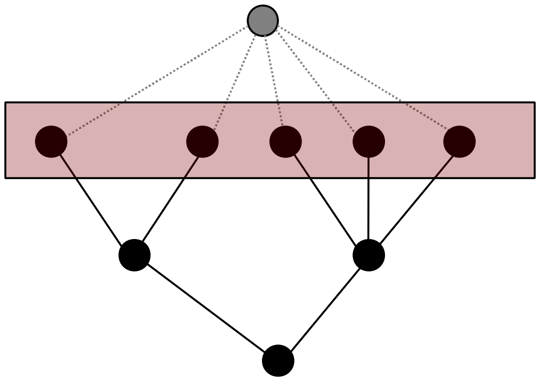
... Mais (à l'issu des dernières journées), quelques inconnues sur les interactions graphes / treillis.



systèmes de classes (atomiques)



Dual co-atomiques



-- idée --

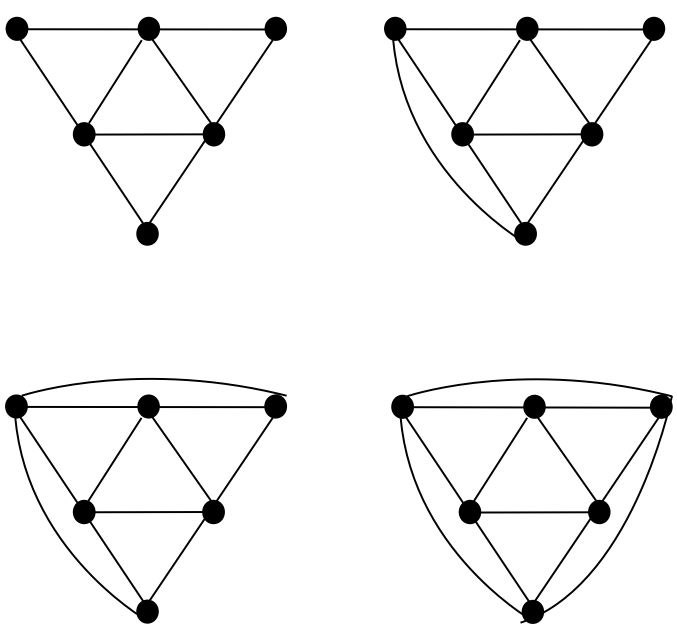
considérer que les co-atomes sont les cliques maximales d'un graphe.

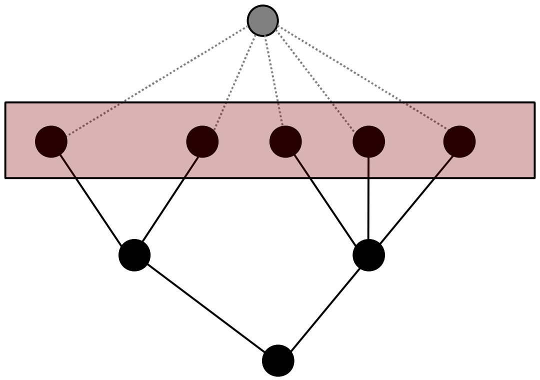
les **quasi-hiérarchies**

sont en bijection

avec les **treillis des cliques maximales**

d'un graphe ne possédant aucune des quatre graphes ci-contre comme sous graphe induit





Et pour les systèmes de classes parcimonieux ?
(treillis démantelables co-atomique)

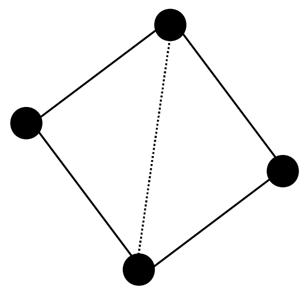
- Un treillis est démantelable ssi p il existe une séquence de *n* sous-treillis de T, $\emptyset = S_0 \subset \ldots \subset S_n = T$ de sous-treillis de T tq $|S_i|=|S_{i-1}|+1$ pour tout i=1,2,...,n

- i.e on peut aboutir à un treillis vidé de tous ses éléments en retirant, à chaque étape, un **élément doublement irréductible**.

parallèle avec les graphes...

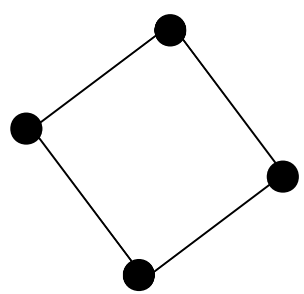
Dans les graphes triangulés, on peut aboutir à un graphe vidé de tous ses sommets en retirant, à chaque étape, **un sommet simplicial.**

Graphes triangulés (cordés)

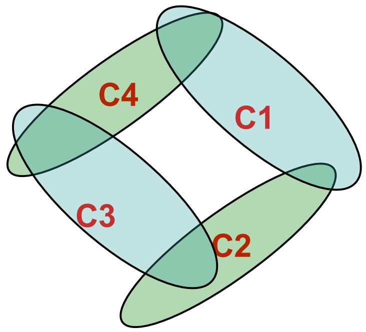


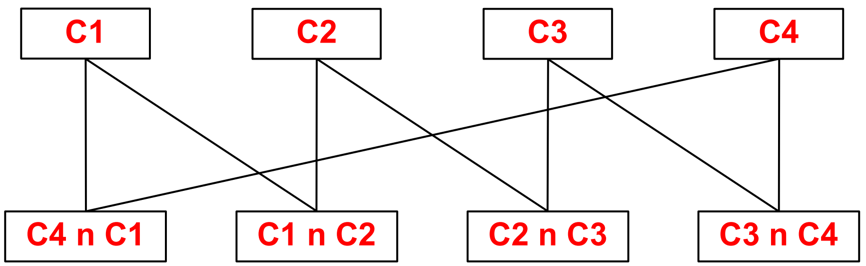
G=(V,E) est un **graphe cordé** ssi tout cycle de sommets de longueur supérieure à 3 est cordé.

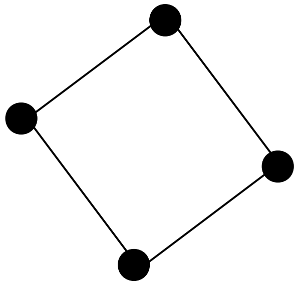
un sommet simplicial est un sommet dont le voisinage est une clique.

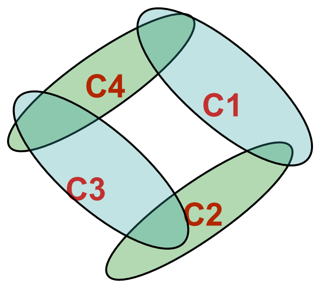


G=(V,E) **graphe cordé** est une **condition nécessaire** à l'obtention d'un treillis des cliques maximales démantelable.

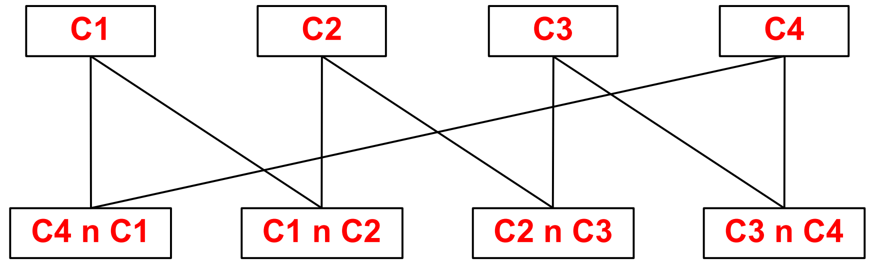




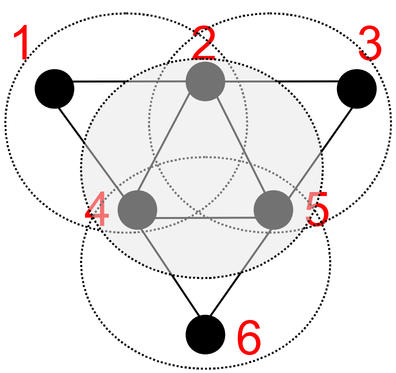


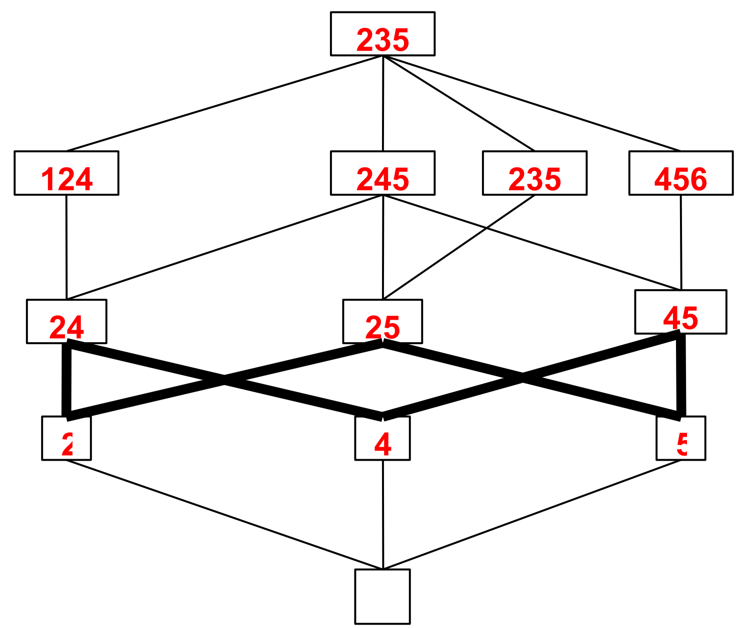


G=(V,E) **graphe cordé**
est une **condition nécessaire** à l'obtention d'un treillis des cliques maximales démantelable.



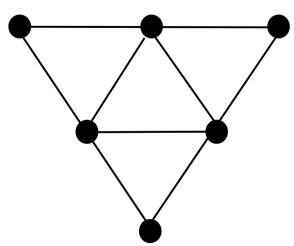
G=(V,E) **graphe cordé**
n'est **pas** une **condition suffisante**



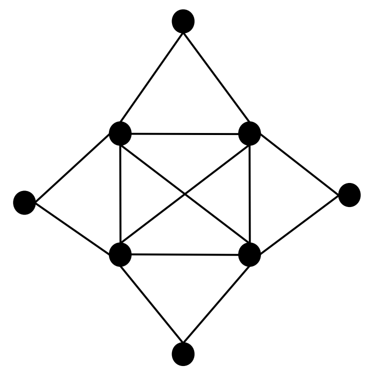


graphes fortement cordés $\subset$ graphes cordés

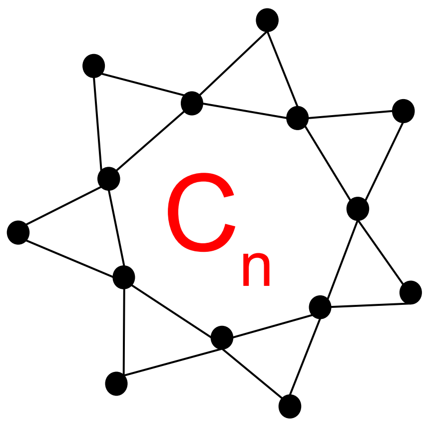
G=(V,E) **graphe fortement cordé** (strongly chordal)
ssi G est **cordé** et **sans soleils induits**.



3-soleil



4-soleil



n-soleil

G=(V,E) **graphe fortement cordé** est caractérisé par

- l'existence d'un schéma d'**élimination fortement simplicial**
- l'existence d'un schéma d'**élimination simple**.

Résultats.

- Les treillis démantelables co-atomiques sont en bijection avec les treillis des cliques maximales d'un graphe fortement cordé G.

- Une feuille, telle que définie dans les systèmes de classes parcimonieux, est un sommet simple de G, tels que définis par Farber en 83

Perspectives & axes de recherches.

- Transfert de connaissances graphes / treillis / classification
    - Visualisation du système de classes (treillis, arbre des cliques, ...)
    - caractérisation de graphes par leur treillis des cliques maximales.

- Plongements de données quelconques dans le modèle
    - propriétés du plongement
    - algorithmique efficace

- Applications sur données réelles

Merci

Questions ?