Densité, VC-dimension et étiquetages de graphes

Soutenance de thèse

Sébastien Ratel

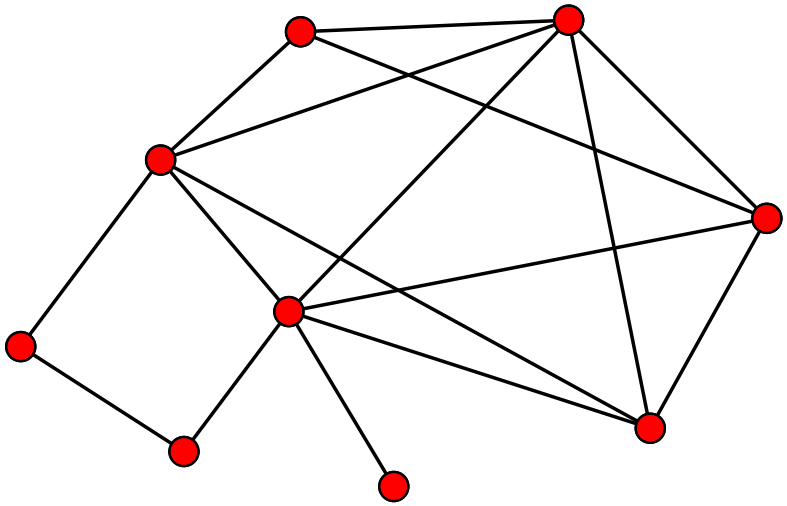
sous la direction de Victor Chepoi et Arnaud Labourel

8 Novembre 2019

LIS
LABORATOIRE D'INFORMATIQUE & SYSTÈMES

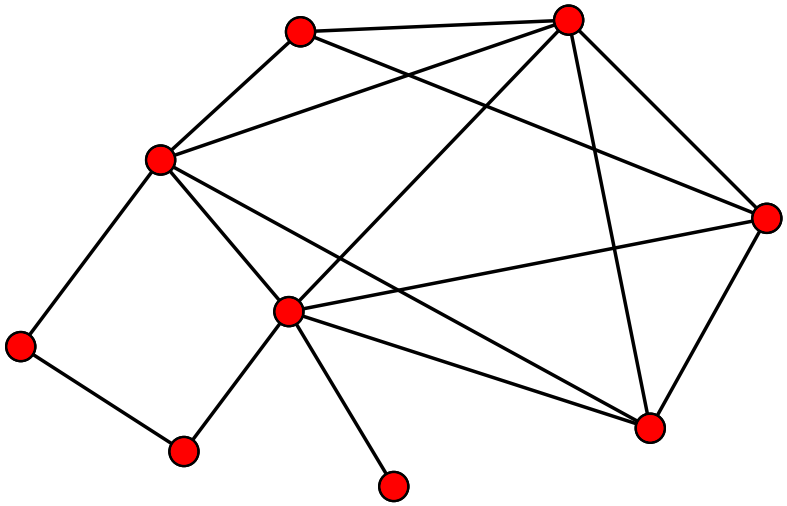
Aix*Marseille université
Initiative d'excellence

Graphes

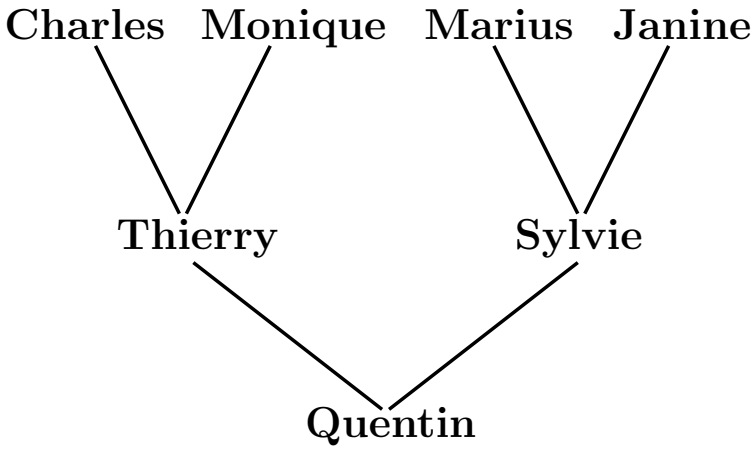


Graphe $G = (V, E)$

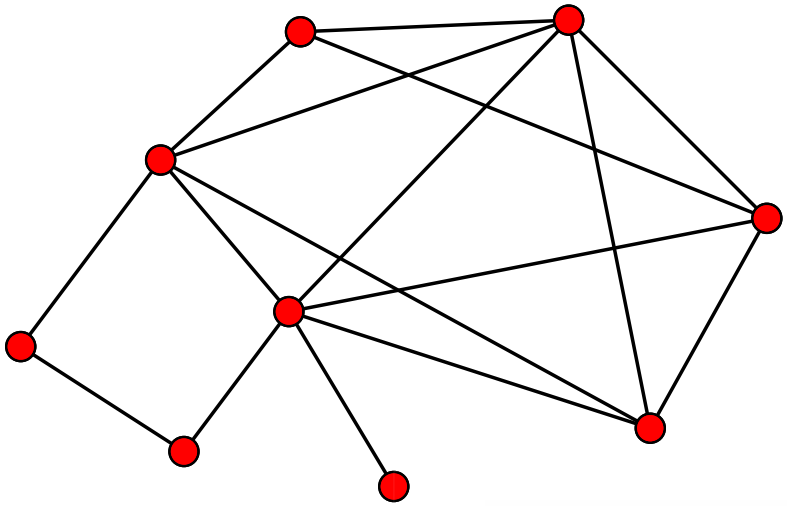
Graphes



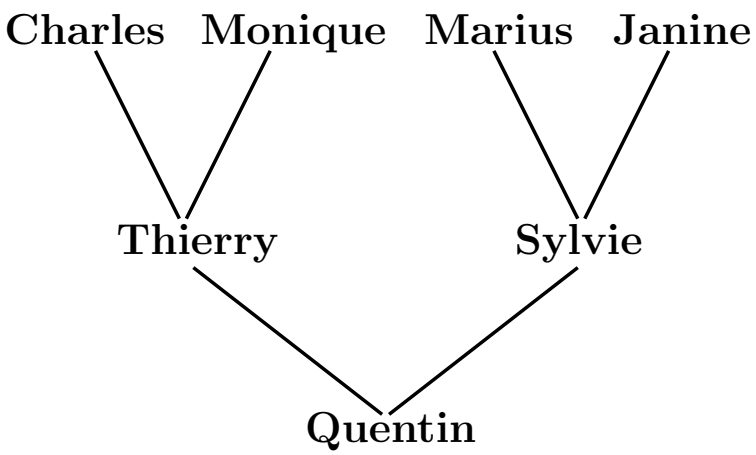
Graphe $G = (V, E)$

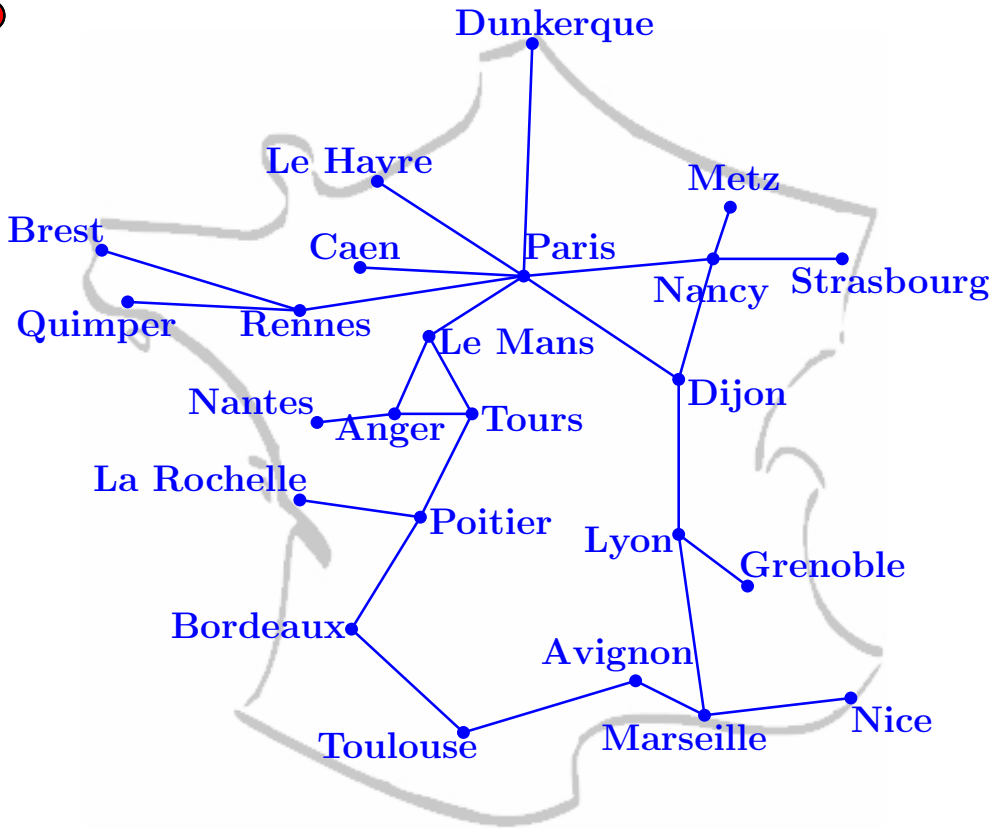


Graphes

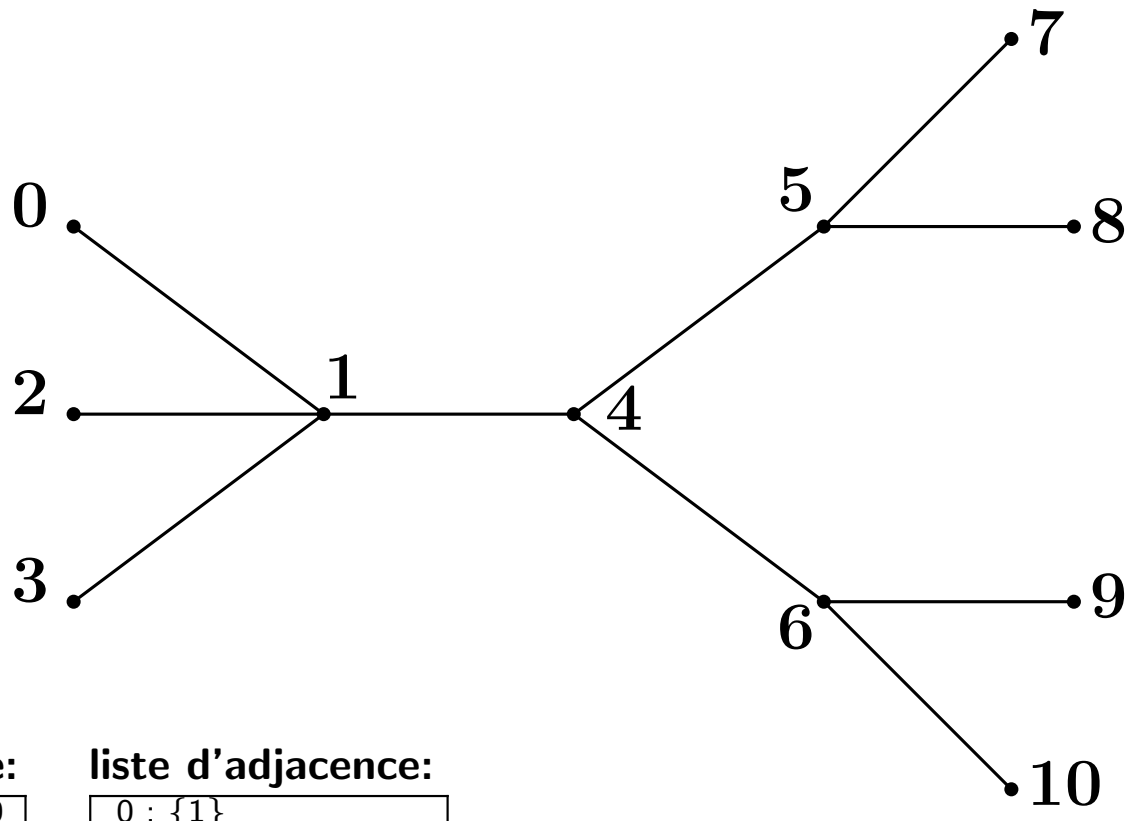


Graphe $G = (V, E)$





Représentations de graphes



matrice d'adjacence:

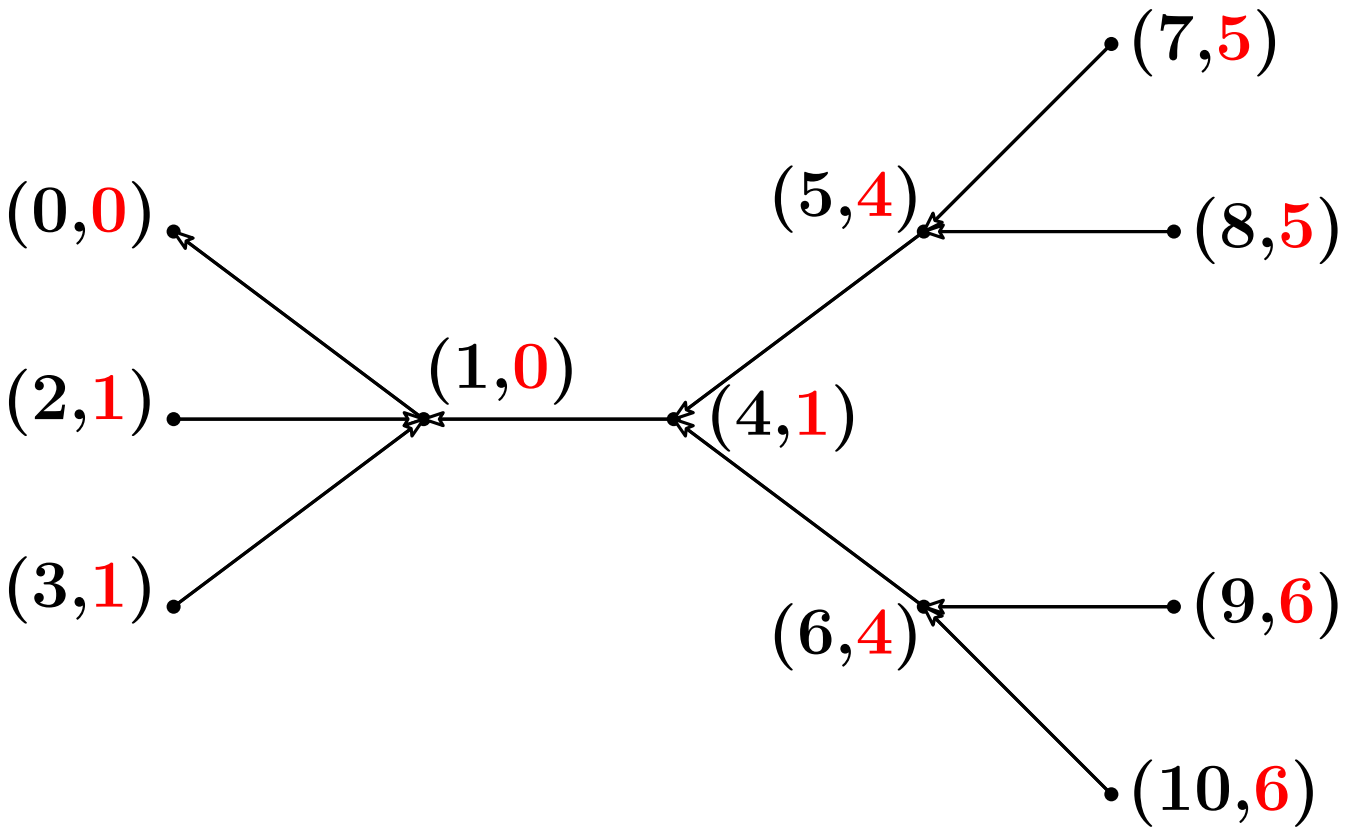
0	1	0	0	0	0	0	0	0	0	0
1	0	1	1	1	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0
0	1	0	0	0	1	1	0	0	0	0
0	0	0	0	1	0	0	1	1	0	0
0	0	0	0	0	0	0	0	0	1	1
0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0

liste d'adjacence:

$0 : \{1\}$
$1 : \{0, 2, 3, 4\}$
$2 : \{1\}$
$3 : \{1\}$
$4 : \{1, 5, 6\}$
$5 : \{4, 7, 8\}$
$6 : \{4, 9, 10\}$
$7 : \{5\}$
$8 : \{5\}$
$9 : \{6\}$
$10 : \{6\}$

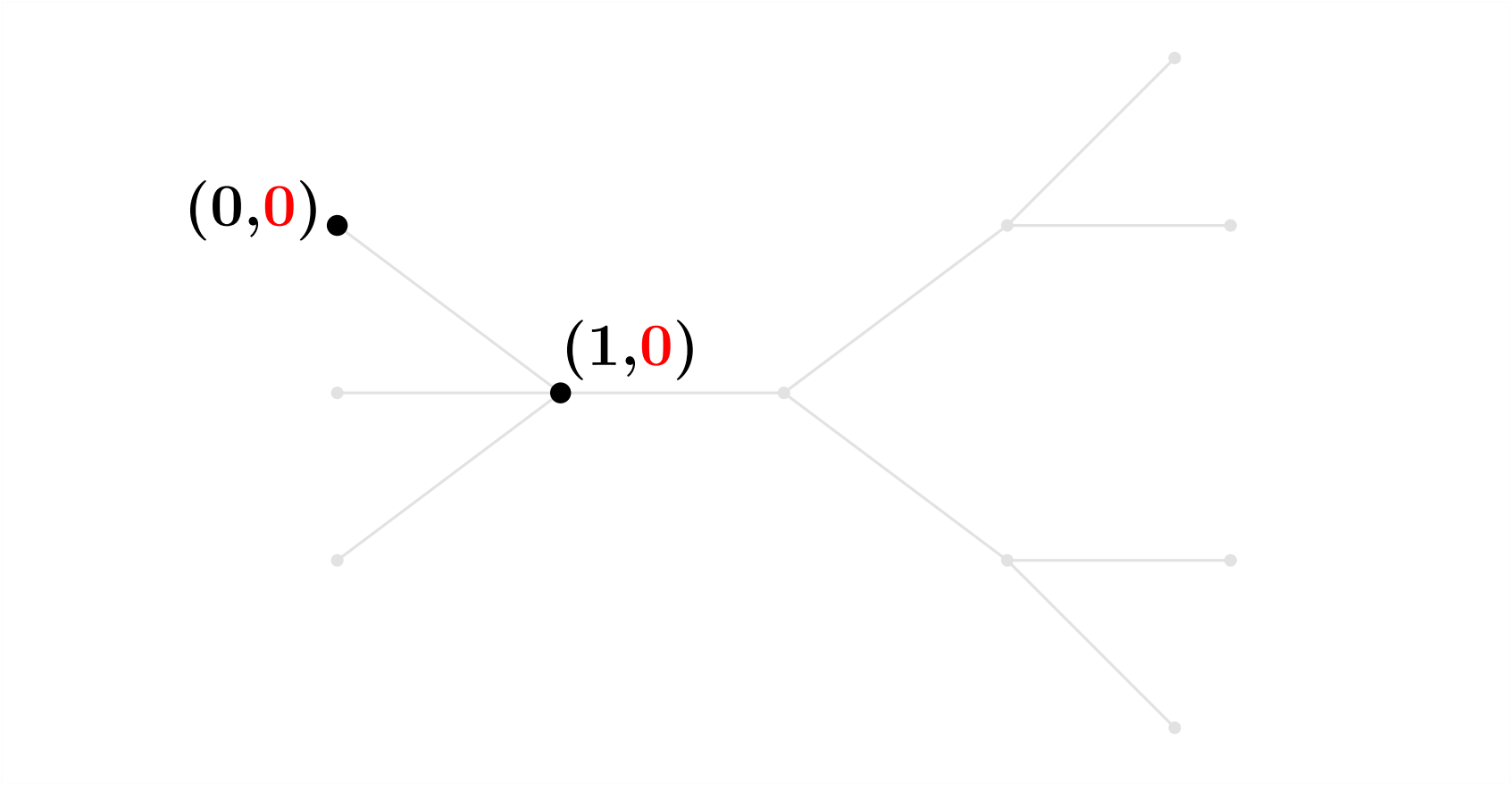
...

Représentations de graphes

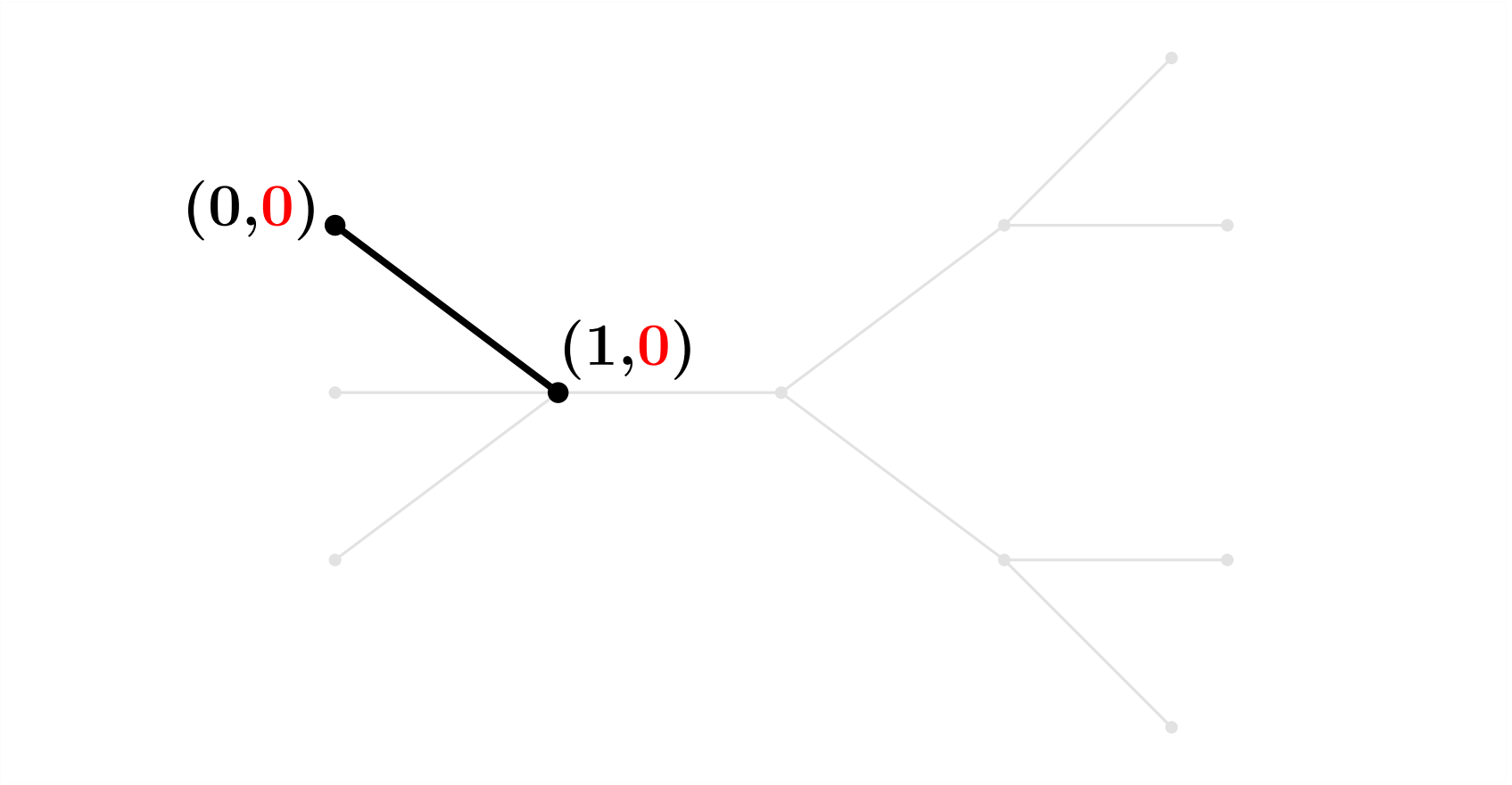


Représentations de graphes

Représentations de graphes



Représentations de graphes



Schémas d'étiquetage

Un **schéma d'étiquetage** sur une famille de graphes $\mathcal{G}$:

- Une **fonction d'encodage** qui :
  - possède la **connaissance globale** d'un graphe $G \in \mathcal{G}$;
  - donne une étiquette à chaque sommet de G;
- Une **fonction de décodage** qui :
  - possède la **connaissance locale** donnée par 2 étiquettes;
  - répond à un type de requête prédéfini : **adjacence** (ALS), **distance** (DLS), **routage** (RLS), ...

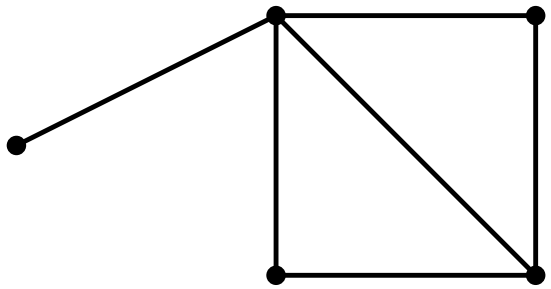
Nous souhaitons que les étiquettes soient :

1. **petites** : de longueur (poly-)logarithmique;
2. **faciles à décoder** : en temps (poly-)logarithmique;
3. **attribuées en temps raisonnable** : polynomial.

Densité et schémas d'adjacence

La **densité** d'un graphe $G = (V, E)$ est $\max_{H \subseteq G} \frac{|E(H)|}{|V(H)|}$.

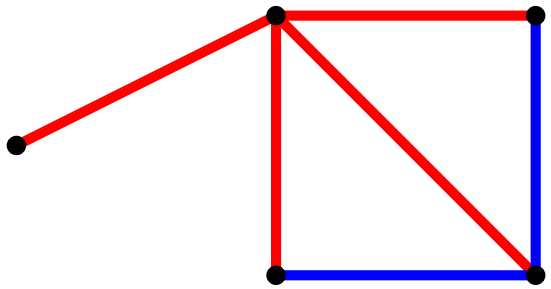
Théorème *(Nash-Williams 1961; Kannan, Naor & Rudich 1992)*. La famille des graphes de densité k sur n sommets admet un ALS utilisant des étiquettes de $O(k \log n)$ bits.



Densité et schémas d'adjacence

La **densité** d'un graphe $G = (V, E)$ est $\max_{H \subseteq G} \frac{|E(H)|}{|V(H)|}$.

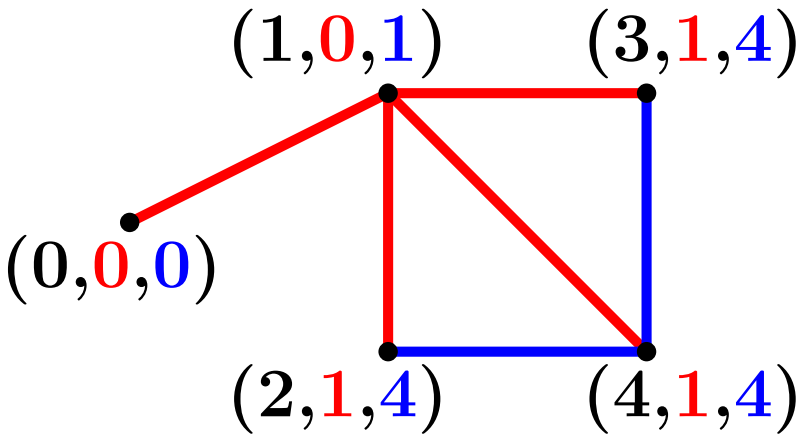
Théorème *(Nash-Williams 1961; Kannan, Naor & Rudich 1992)*. La famille des graphes de densité k sur n sommets admet un ALS utilisant des étiquettes de $O(k \log n)$ bits.



Densité et schémas d'adjacence

La **densité** d'un graphe $G = (V, E)$ est $\max_{H \subseteq G} \frac{|E(H)|}{|V(H)|}$.

Théorème *(Nash-Williams 1961; Kannan, Naor & Rudich 1992)*. La famille des graphes de densité k sur n sommets admet un ALS utilisant des étiquettes de $O(k \log n)$ bits.



Contributions de la thèse

I. Densité et VC-dimension

Densité des sous-graphes de produits cartésiens

Densité des sous-graphes de demi-cubes

II. Schémas d'étiquetages de distance

Schéma de distance pour les graphes médians sans Q_3

Schéma de distance pour les graphes pontés sans K_4

Densité et VC-dimension

Densité des sous-graphes d'hypercubes

Théorème *(Haussler, Littlestone & Warmuth 1994)*.
Si $G(\mathcal{S}) = (\mathcal{S}, E)$ est un sous-graphe d'hypercube, alors

$$\frac{|E|}{|\mathcal{S}|} \leq \text{VC-dim}(\mathcal{S}) \leq \log |\mathcal{S}|.$$

Corollaire.
Les sous-graphes d'hypercubes n sommets et de VC-dimension d admettent un ALS avec étiquettes de $O(d \cdot \log n)$ bits.

Nos résultats

Théorème 1 *(Chepoi, Labourel & R., JGT à paraître)*.
Si $G = (V, E)$ est un sous-graphe d'un produit cartésien de graphes connexes et sans mineur H, alors

$$\frac{|E|}{|V|} \leq c_H \cdot \text{VC-dim}^*(G) \leq c_H \cdot \log |V|.$$

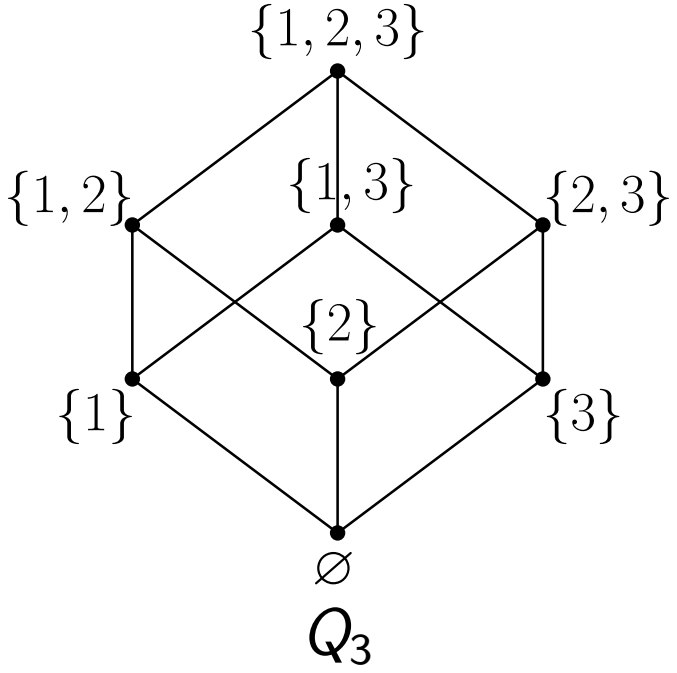
Théorème 2 *(Chepoi, Labourel & R., EJC 2019)*.
Si $G(\mathcal{S}) = (\mathcal{S}, E)$ est un sous-graphe de demi-cube, alors

$$\frac{|E|}{|\mathcal{S}|} \leq \binom{\text{cVC-dim}^*(\mathcal{S})}{2}.$$

Hypercubes

L'**hypercube de dimension** m est le graphe Q_m tel que :

- $V(Q_m) := 2^X$, avec $X := \{1, \ldots, m\}$;
- $AB \in E(Q_m)$ ssi $|A \triangle B| = 1$.

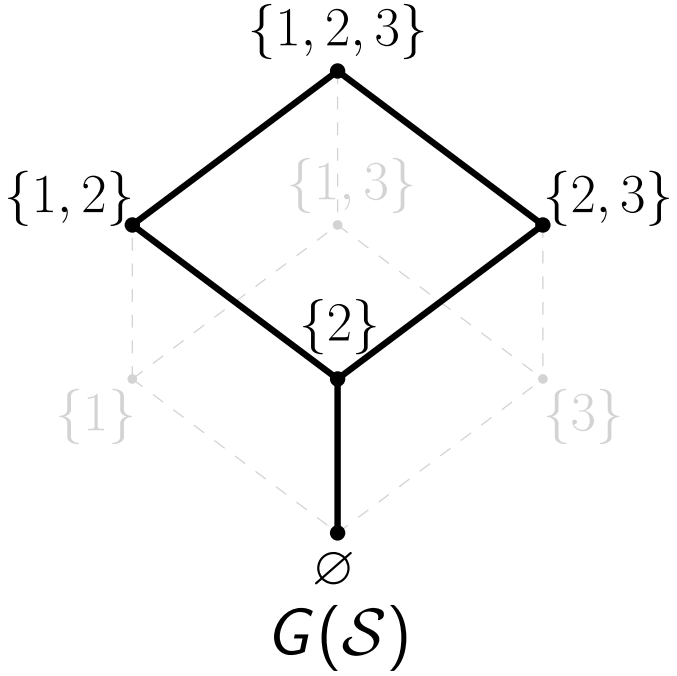


Q_3

Hypercubes

L'**hypercube de dimension** m est le graphe Q_m tel que :

- $V(Q_m) := 2^X$, avec $X := \{1, \ldots, m\}$;
- $AB \in E(Q_m)$ ssi $|A \triangle B| = 1$.



$G(\mathcal{S})$

Dimension de Vapnik-Chervonenkis (1/2)

$\mathcal{S} \subseteq 2^X$, $Y \subseteq X$.

Trace de $\mathcal{S}$ sur Y : $\mathcal{S}_Y := \{S \cap Y : S \in \mathcal{S}\}$.

Y **pulvérisé** par $\mathcal{S}$: $\mathcal{S}_Y = 2^Y$.

VC-dimension VC-dim($\mathcal{S}$) : plus grand Y pulvérisé par $\mathcal{S}$.

Exemple.

$$
\begin{aligned}
X &:= \{1,2,3,4\} \\
Y &:= \{1,3\} \\
\mathcal{S} &= \{\varnothing, \{1\}, \{2\}, \{2,3\}, \{1,2,3\}, \{1,2,4\}\} \\
\mathcal{S}_Y &= \{\varnothing, \{1\}, \varnothing, \ \{3\}, \ \{1,3\}, \ \{1\} \ \}
\end{aligned}
$$

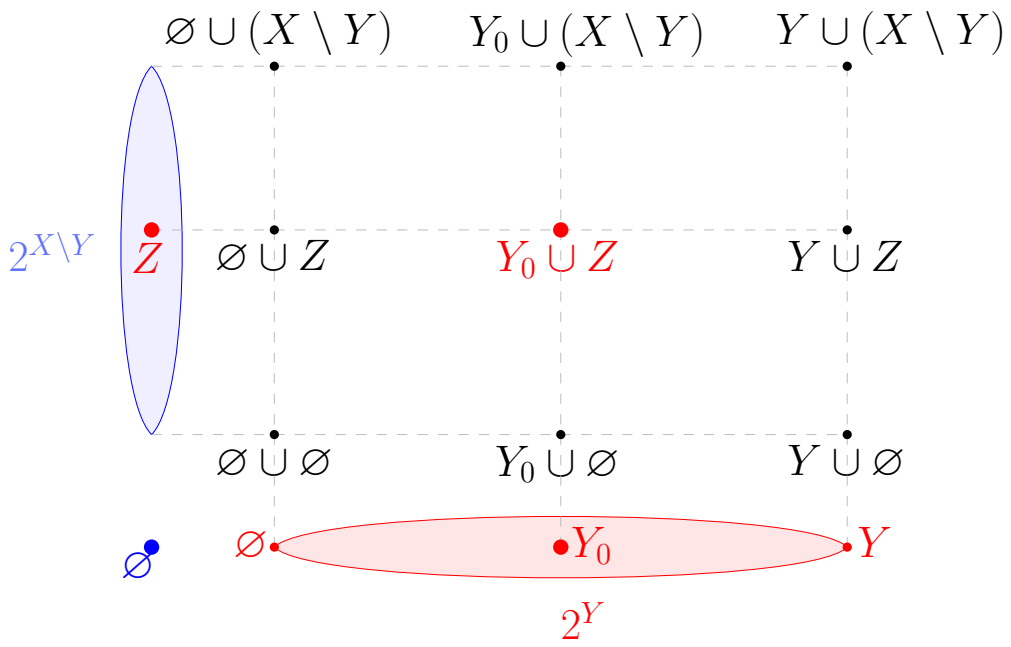
Dimension de Vapnik-Chervonenkis (1/2)

$\mathcal{S} \subseteq 2^X$, $Y \subseteq X$.

Trace de $\mathcal{S}$ sur Y : $\mathcal{S}_Y := \{S \cap Y : S \in \mathcal{S}\}$.

Y **pulvérisé** par $\mathcal{S}$: $\mathcal{S}_Y = 2^Y$.

VC-dimension VC-dim($\mathcal{S}$) : plus grand Y pulvérisé par $\mathcal{S}$.



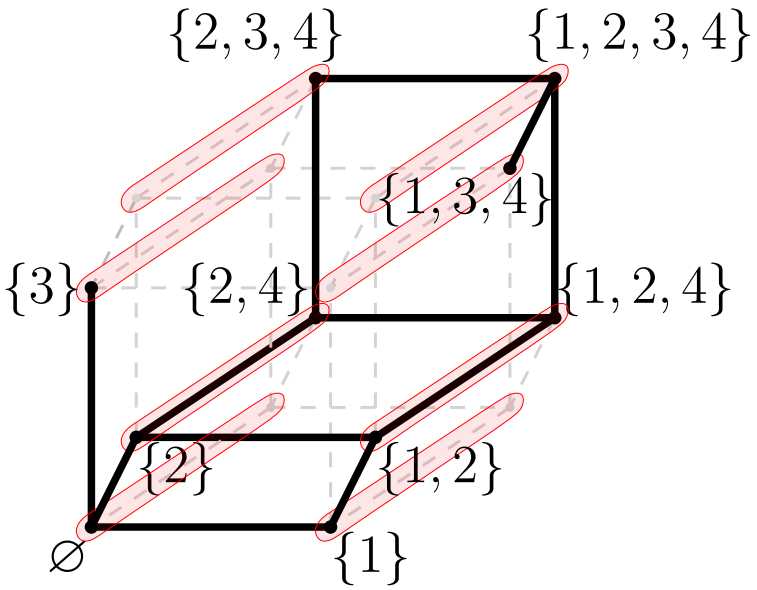
Dimension de Vapnik-Chervonenkis (1/2)

$\mathcal{S} \subseteq 2^X$, $Y \subseteq X$.

Trace de $\mathcal{S}$ sur Y : $\mathcal{S}_Y := \{S \cap Y : S \in \mathcal{S}\}$.

Y **pulvérisé** par $\mathcal{S}$: $\mathcal{S}_Y = 2^Y$.

VC-dimension VC-dim($\mathcal{S}$) : plus grand Y pulvérisé par $\mathcal{S}$.

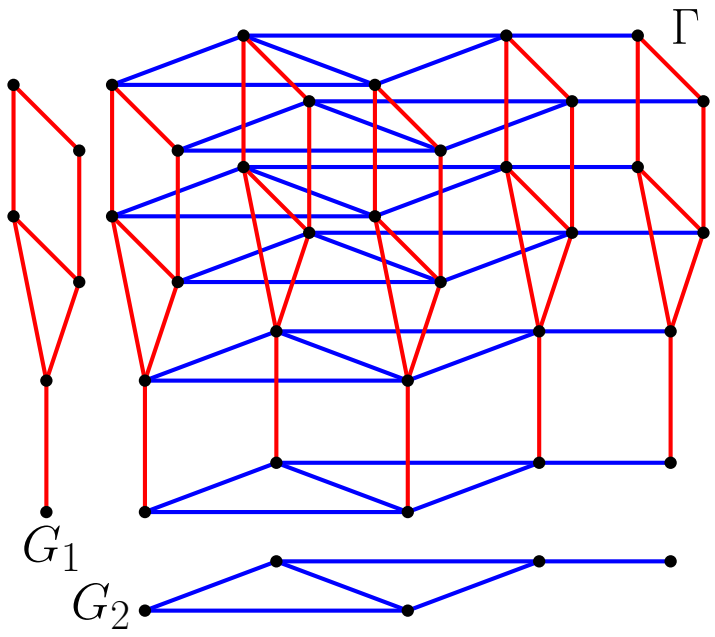


$Y = \{1, 2, 3\}$ pulvérisé
$\Rightarrow$ VC-dim($\mathcal{S}$) ≥ 3

Produit cartésien de graphes

Le **produit cartésien** $\Gamma := G_1 \square \cdots \square G_m$ de m graphes est le graphe tel que :

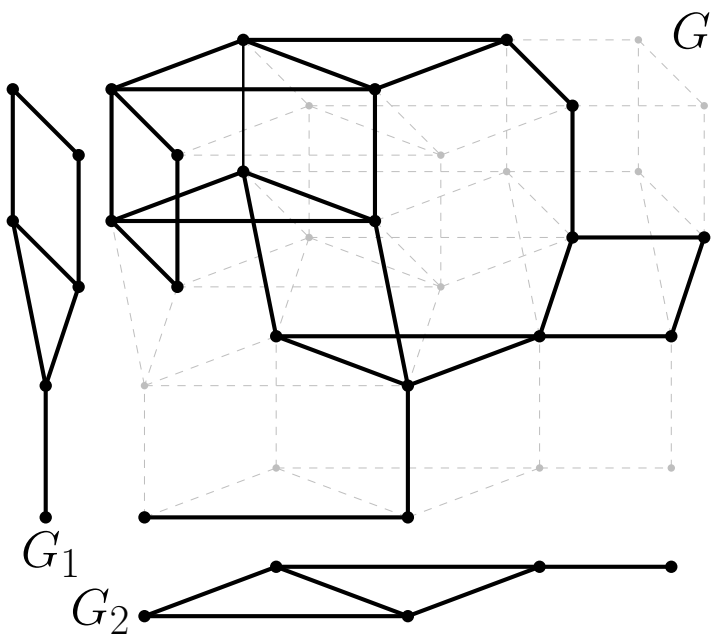
- $V(\Gamma) := V(G_1) \times \cdots \times V(G_m)$;
- $(u_1, \ldots, u_m)(v_1, \ldots, v_m) \in E(\Gamma)$ ssi $u_i v_i \in E(G_i)$ pour un certain i, et $u_j = v_j$ pour tout $j \neq i$.



Produit cartésien de graphes

Le **produit cartésien** $\Gamma := G_1 \square \cdots \square G_m$ de m graphes est le graphe tel que :

- $V(\Gamma) := V(G_1) \times \cdots \times V(G_m)$;
- $(u_1, \ldots, u_m)(v_1, \ldots, v_m) \in E(\Gamma)$ ssi $u_i v_i \in E(G_i)$ pour un certain i, et $u_j = v_j$ pour tout $j \neq i$.

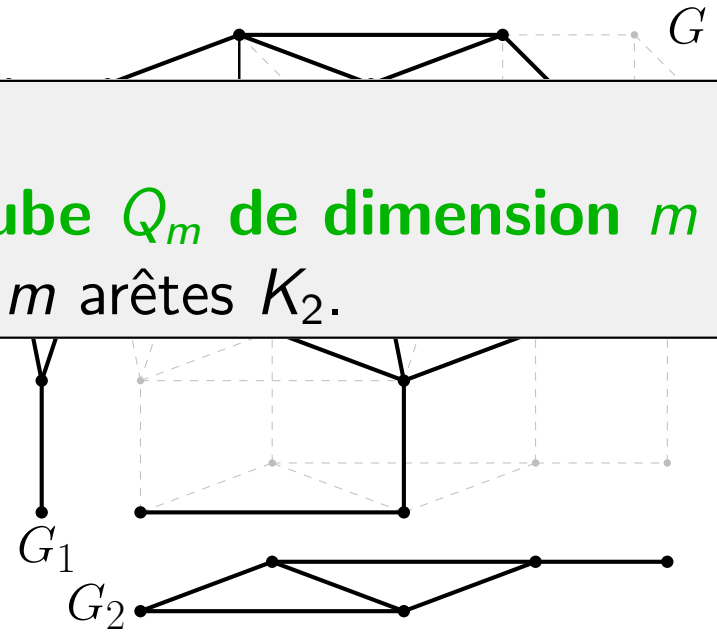


G sous-graphe de Γ

Produit cartésien de graphes

Le **produit cartésien** $\Gamma := G_1 \square \cdots \square G_m$ de m graphes est le graphe tel que :

- $V(\Gamma) := V(G_1) \times \cdots \times V(G_m)$;
- $(u_1, \ldots, u_m)(v_1, \ldots, v_m) \in E(\Gamma)$ ssi $u_i v_i \in E(G_i)$ pour un certain i, et $u_j = v_j$ pour tout $j \neq i$.



G sous-graphe de Γ

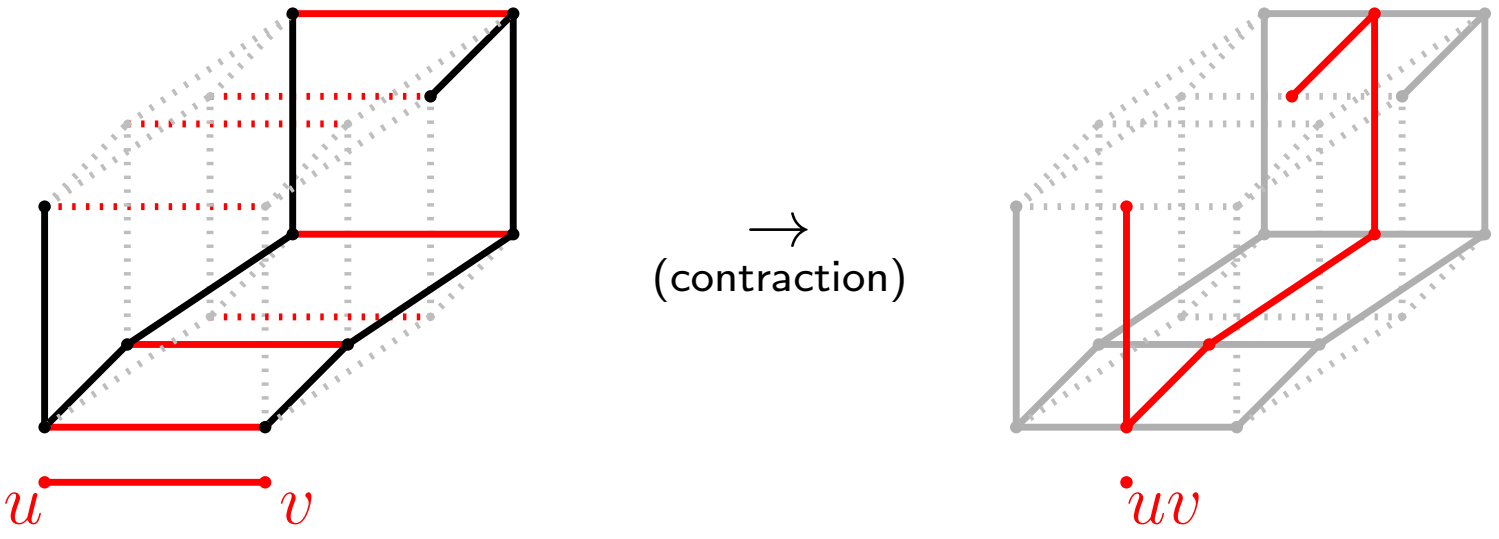
Remarque.

L'**hypercube Q_m de dimension m** est le produit cartésien de m arêtes K_2.

Dimension de Vapnik-Chervonenkis (2/2)

La **VC-dimension** VC-dim(G) de $G \subseteq Q_m$ est la plus grande dimension d d'un hypercube Q_d constructible depuis G via

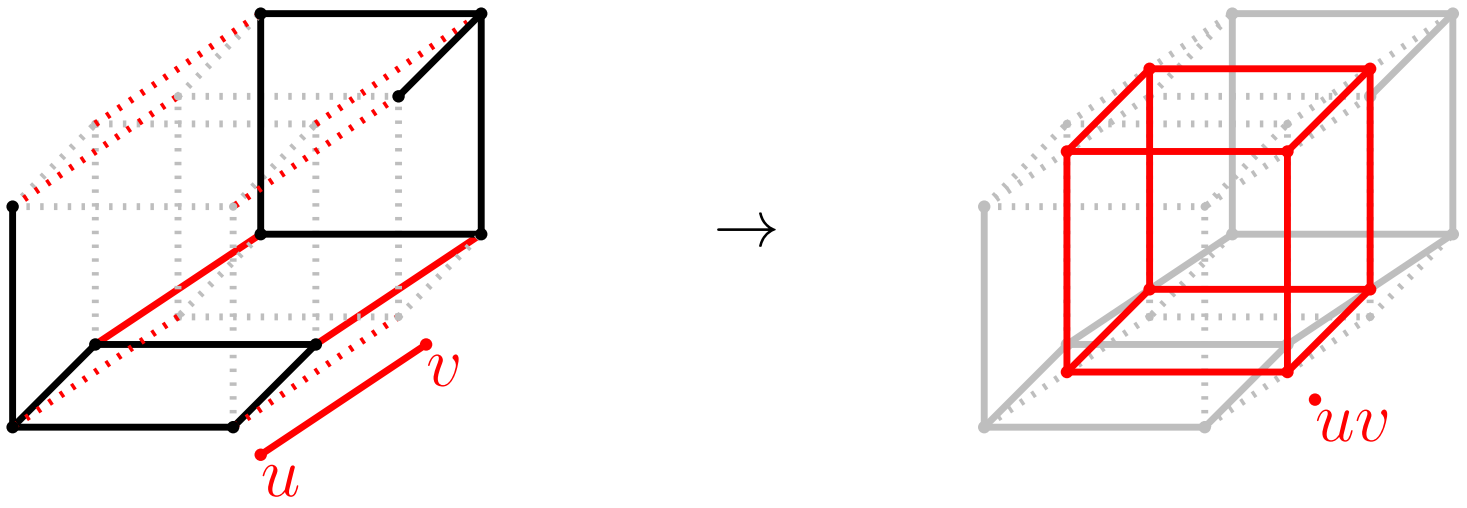
(contraction) choisir un facteur K_2 de Q_m ; **contracter** chaque copie du K_2 dans G ($uv \in V(H) \Leftrightarrow u \in V$ **ou** $v \in V$).



Dimension de Vapnik-Chervonenkis (2/2)

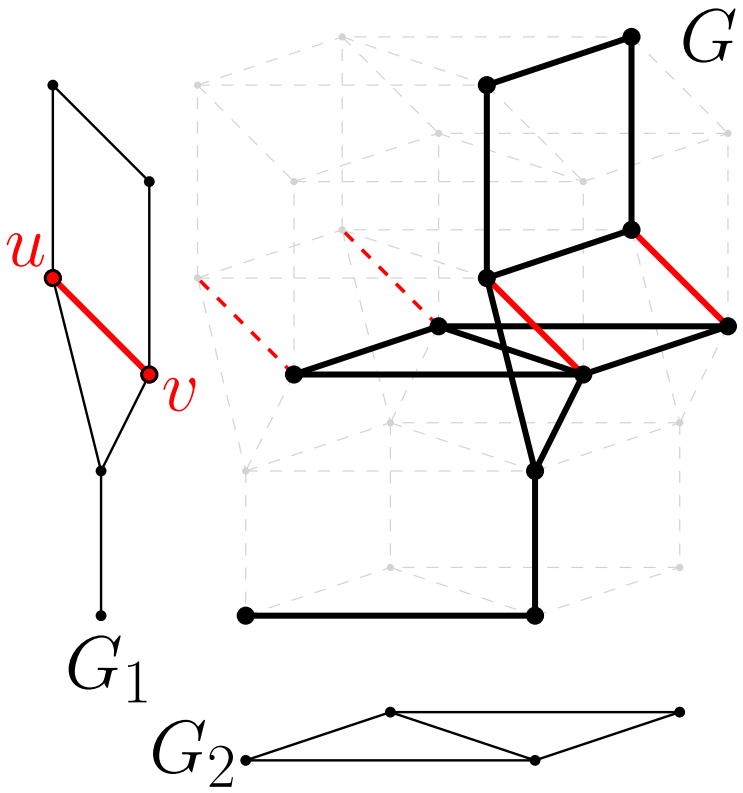
La **VC-dimension** VC-dim(G) de $G \subseteq Q_m$ est la plus grande dimension d d'un hypercube Q_d constructible depuis G via

(contraction) choisir un facteur K_2 de Q_m ; **contracter** chaque copie du K_2 dans G ($uv \in V(H) \Leftrightarrow u \in V$ **ou** $v \in V$).



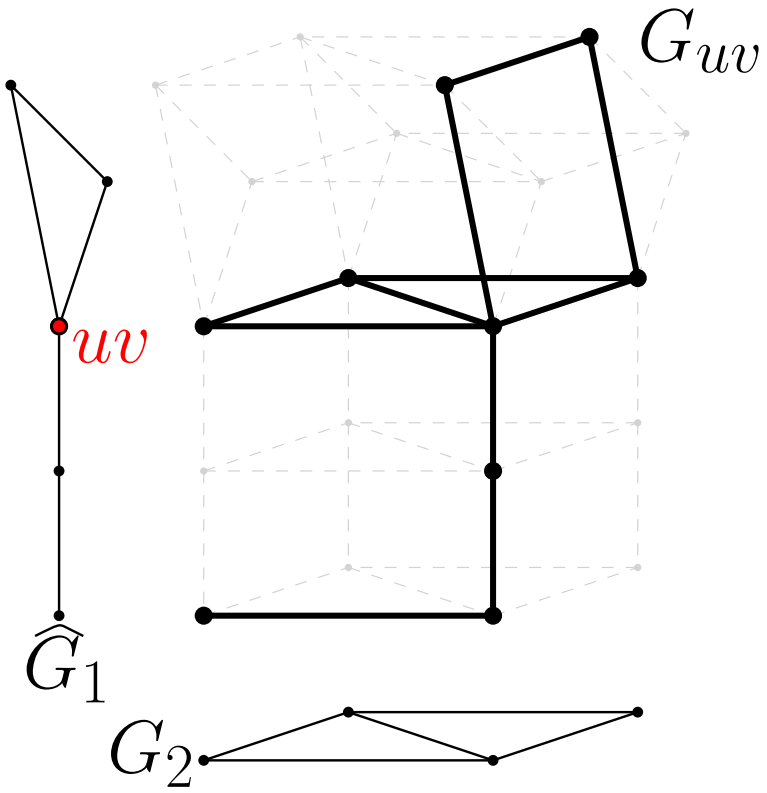
Densité des sous-graphes de produits cartésiens

Schéma de preuve du théorème 1: $\frac{|E|}{|V|} \leq c_H \cdot \text{VC-dim}^*(G)$.



Densité des sous-graphes de produits cartésiens

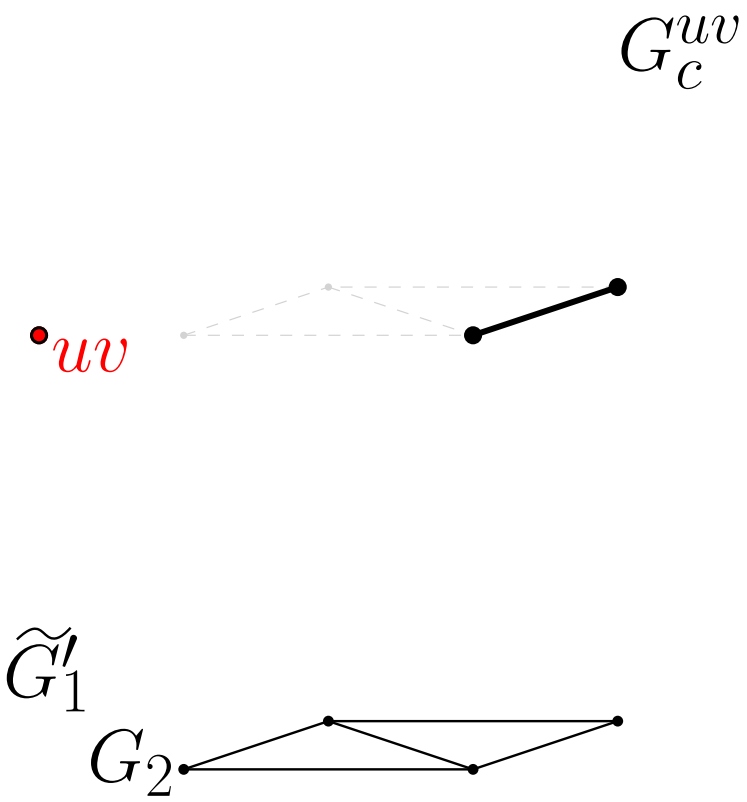
Schéma de preuve du théorème 1: $\frac{|E|}{|V|} \leq c_H \cdot \text{VC-dim}^*(G)$.



Lemme 1. $\text{VC-dim}^*(G_{uv}) \leq \text{VC-dim}^*(G)$

Densité des sous-graphes de produits cartésiens

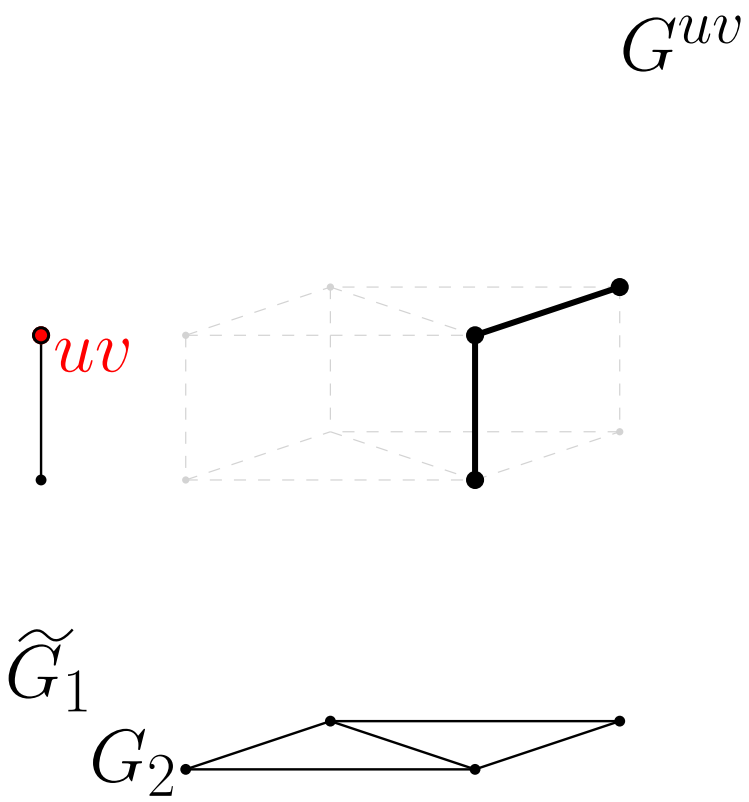
Schéma de preuve du théorème 1: $\frac{|E|}{|V|} \leq c_H \cdot \text{VC-dim}^*(G)$.



Lemme 2. $\text{VC-dim}^*(G_c^{uv}) \leq \text{VC-dim}^*(G) - 1$

Densité des sous-graphes de produits cartésiens

Schéma de preuve du théorème 1: $\frac{|E|}{|V|} \leq c_H \cdot \text{VC-dim}^*(G)$.

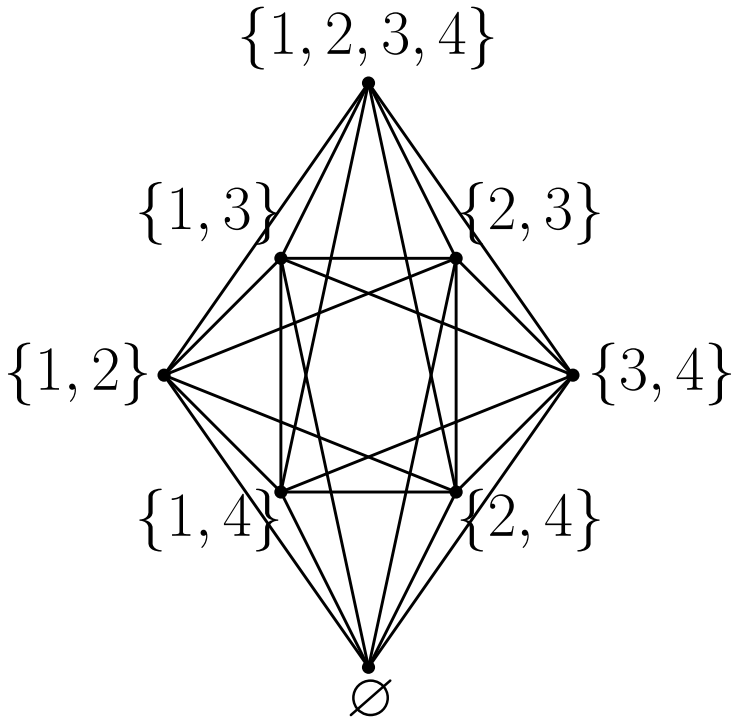


Lemme 3. $\begin{cases} |V(G)| & = & |V(G_{uv})| + |V(G_c^{uv})|, \\ |E(G)| & \leq & |E(G_{uv})| + |E(G_c^{uv})| + |V(G^{uv})|. \end{cases}$

Demi-cubes

Le **demi-cube de dimension** m est le graphe $\frac{1}{2}Q_m$ tel que:

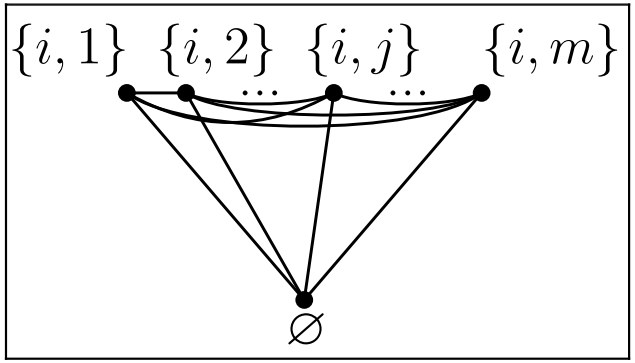
- $V(\frac{1}{2}Q_m) := \{A \subseteq X = \{1, \ldots, m\} : |A| \text{ pair}\}$;
- $AB \in E(\frac{1}{2}Q_m)$ ssi $|A \triangle B| = 2$.

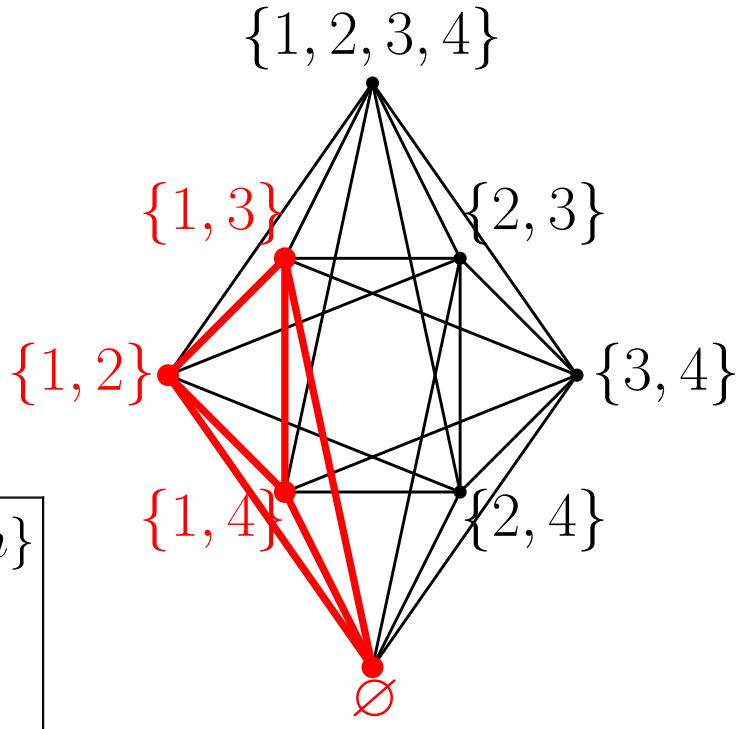


Demi-cubes

Le **demi-cube de dimension** m est le graphe $\frac{1}{2}Q_m$ tel que:

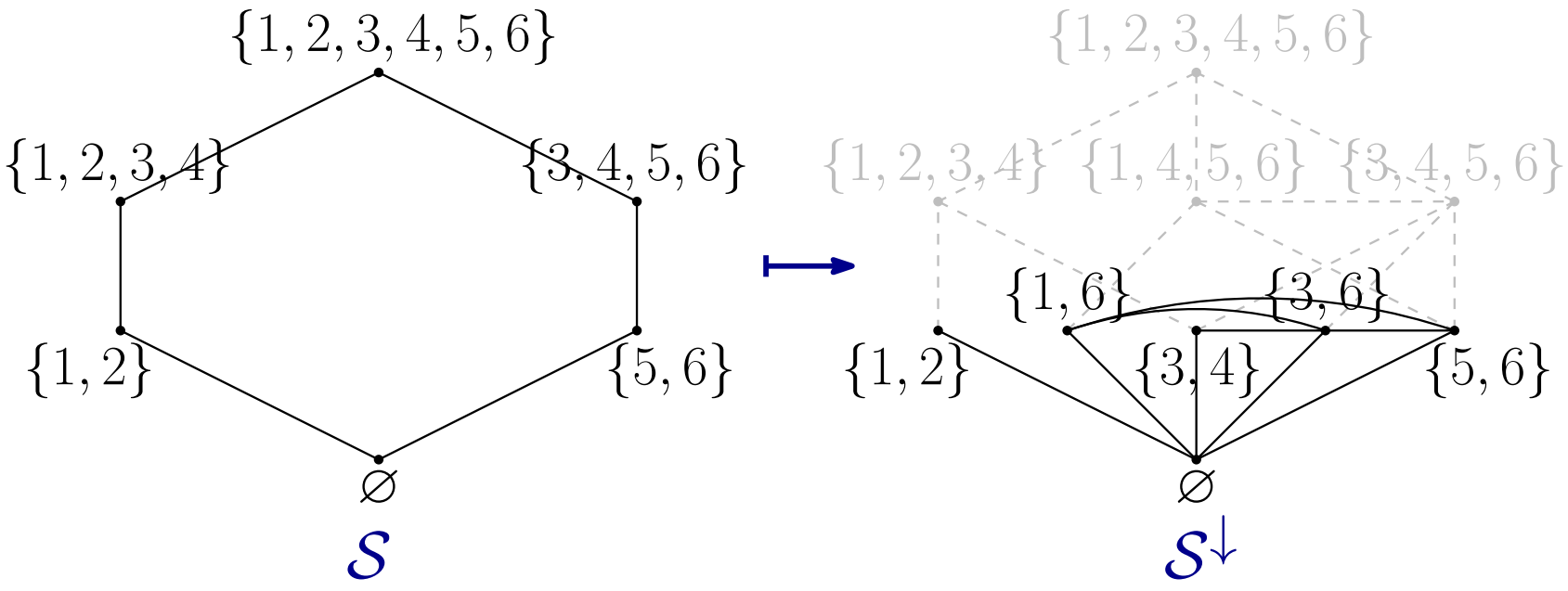
- $V(\frac{1}{2}Q_m) := \{A \subseteq X = \{1, \ldots, m\} : |A| \text{ pair}\}$;
- $AB \in E(\frac{1}{2}Q_m)$ ssi $|A \triangle B| = 2$.





Densité des sous-graphes de demi-cubes

Schéma de preuve du théorème 2 : $\frac{|E|}{|\mathcal{S}|} \leq \binom{\text{cVC-dim}^*(\mathcal{S})}{2}$.



Définir une opération de **d-décalage** de $\mathcal{S}$ en $\mathcal{S}^{\downarrow}$ telle que

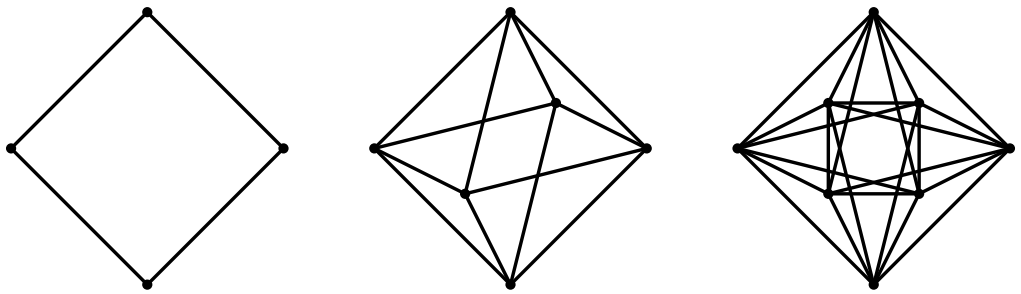
$$\frac{|E|}{|\mathcal{S}|} \leq \frac{|E^{\downarrow}|}{|\mathcal{S}^{\downarrow}|} \leq \binom{\text{cVC-dim}^*(\mathcal{S}^{\downarrow})}{2} \leq \binom{\text{cVC-dim}^*(\mathcal{S})}{2}$$

Corollaires

Pour $G = (V, E) \subseteq G_1 \square \cdots \square G_m$, un ALS attribue des étiquettes de longueur:

- $O(c_H \cdot \text{VC-dim}^*(G) \cdot \log |V|)$, pour G_i **sans mineur** H;
- $O(\text{VC-dim}^*(G) \cdot \log |V|)$, pour G_i **arêtes**;
- $O(\omega(G) \cdot \text{VC-dim}^*(G) \cdot \log |V|)$, pour G_i **démontables**;
- $O(\omega(G) \cdot \text{VC-dim}^*(G) \cdot \log |V|)$, pour G_i **octaèdres**.

Pour $G(\mathcal{S})$ sous-graphe de **demi-cube**, ces étiquettes sont de longueur $O((\text{cVC-dim}^*(\mathcal{S}))^2 \cdot \log |\mathcal{S}|)$.

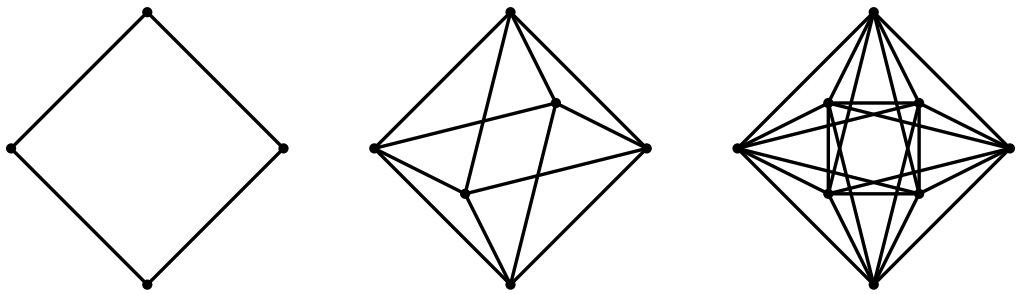


Corollaires

Pour $G = (V, E) \subseteq G_1 \square \cdots \square G_m$, un ALS attribue des étiquettes de longueur:

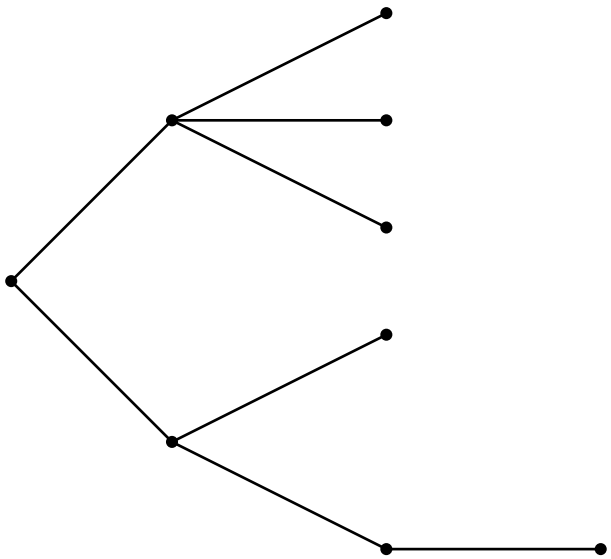
- $O(c_H \cdot \text{VC-dim}^*(G) \cdot \log |V|)$, pour G_i **sans mineur** H;
- $O(\text{VC-dim}^*(G) \cdot \log |V|)$, pour G_i **arêtes**;
- $O(\omega(G) \cdot \text{VC-dim}^*(G) \cdot \log |V|)$, pour G_i **démontables**;
- $O(\omega(G) \cdot \text{VC-dim}^*(G) \cdot \log |V|)$, pour G_i **octaèdres**.

Pour $G(\mathcal{S})$ sous-graphe de **demi-cube**, ces étiquettes sont de longueur $O((\text{cVC-dim}^*(\mathcal{S}))^2 \cdot \log |\mathcal{S}|)$.

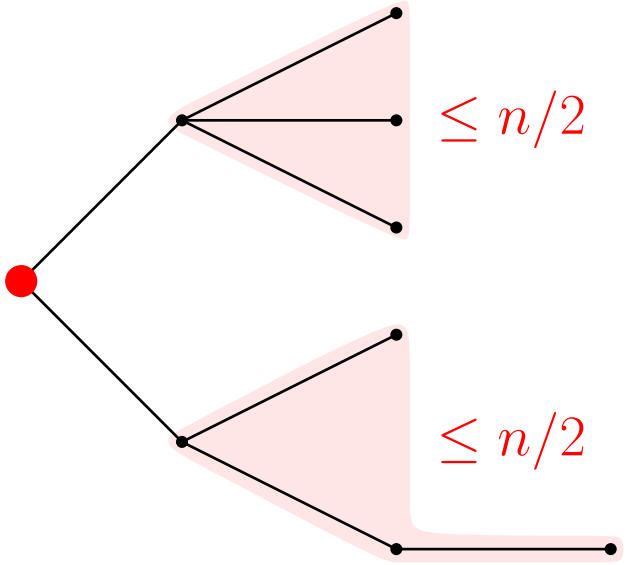


Schémas d'étiquetages de distance (DLS)

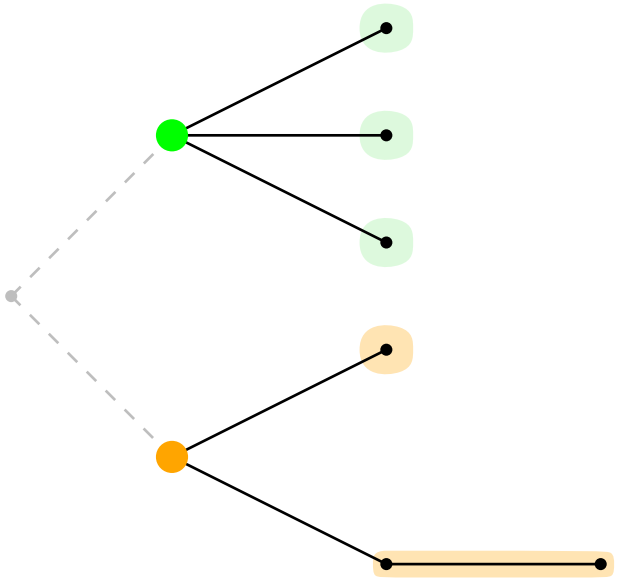
DLS pour les arbres : idée



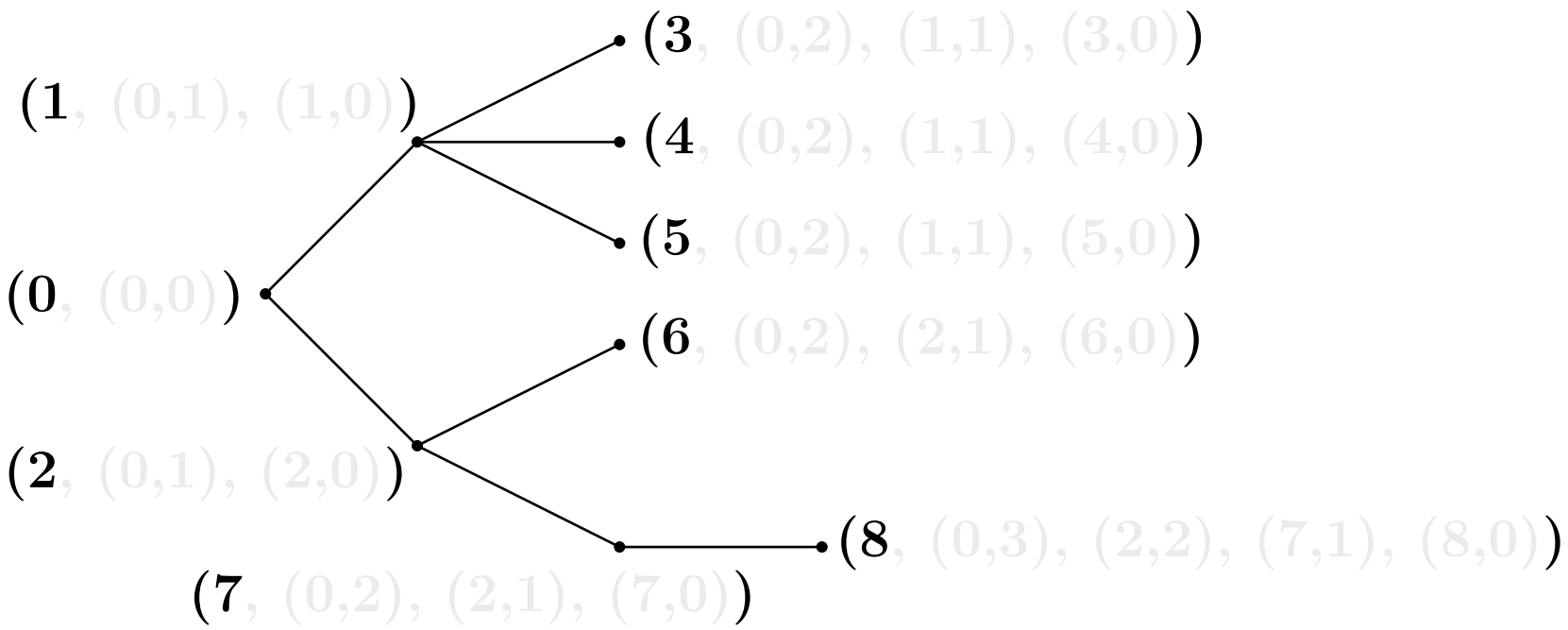
DLS pour les arbres : idée



DLS pour les arbres : idée

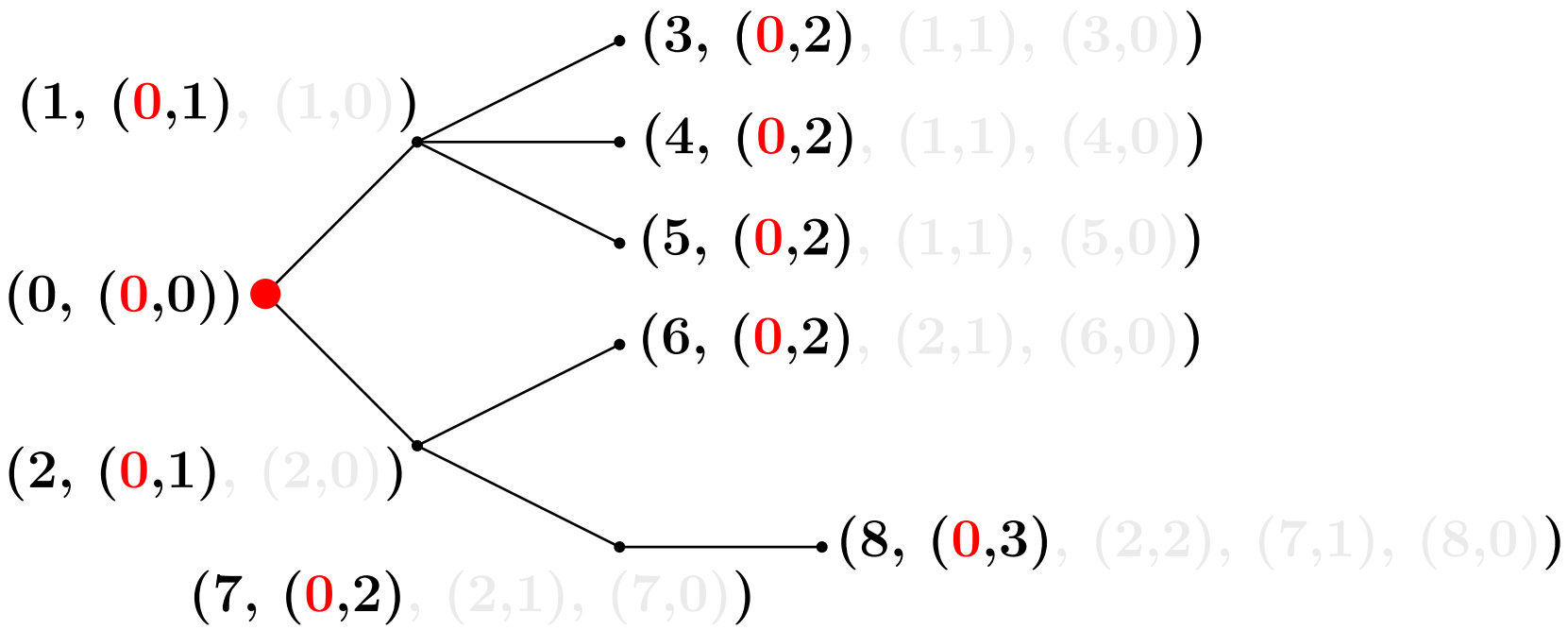


DLS pour les arbres : encodage



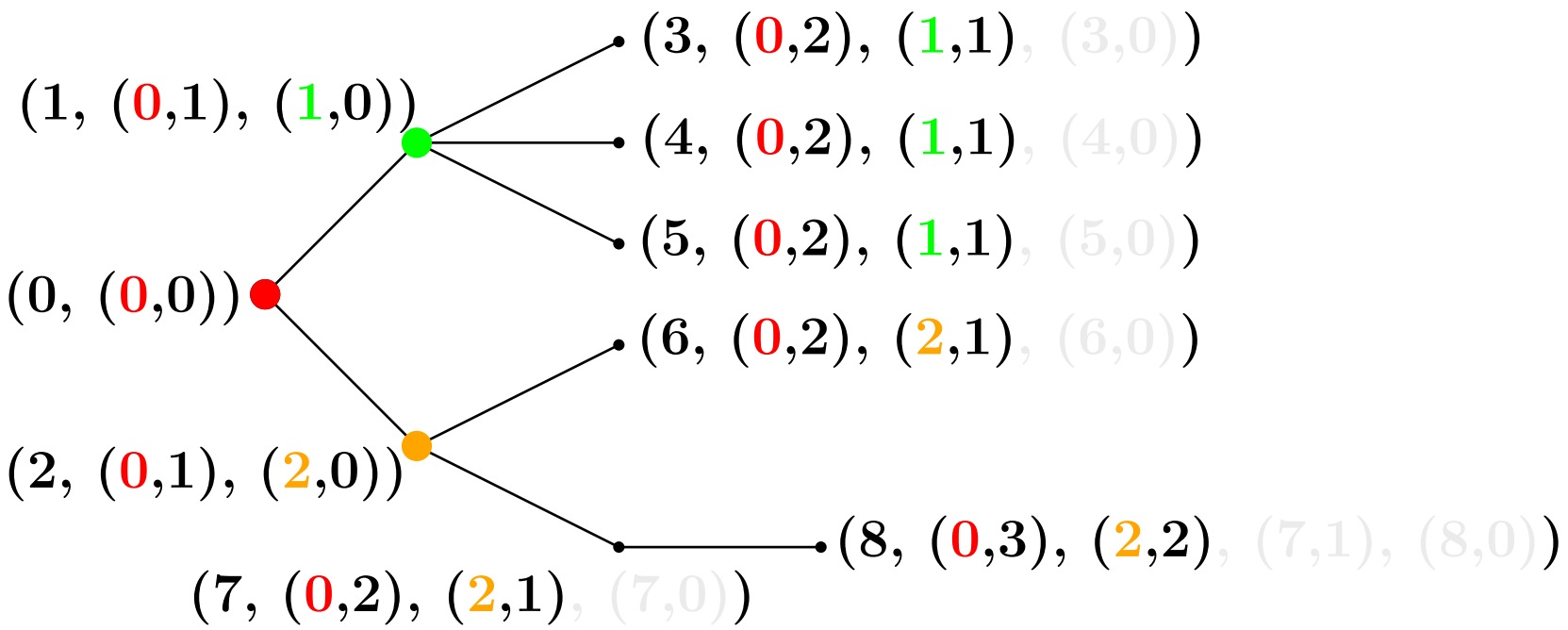
Théorème *(Peleg 2000)*. La famille des arbres sur n sommets admet un DLS avec étiquettes de $O(\log^2 n)$ bits.

DLS pour les arbres : encodage



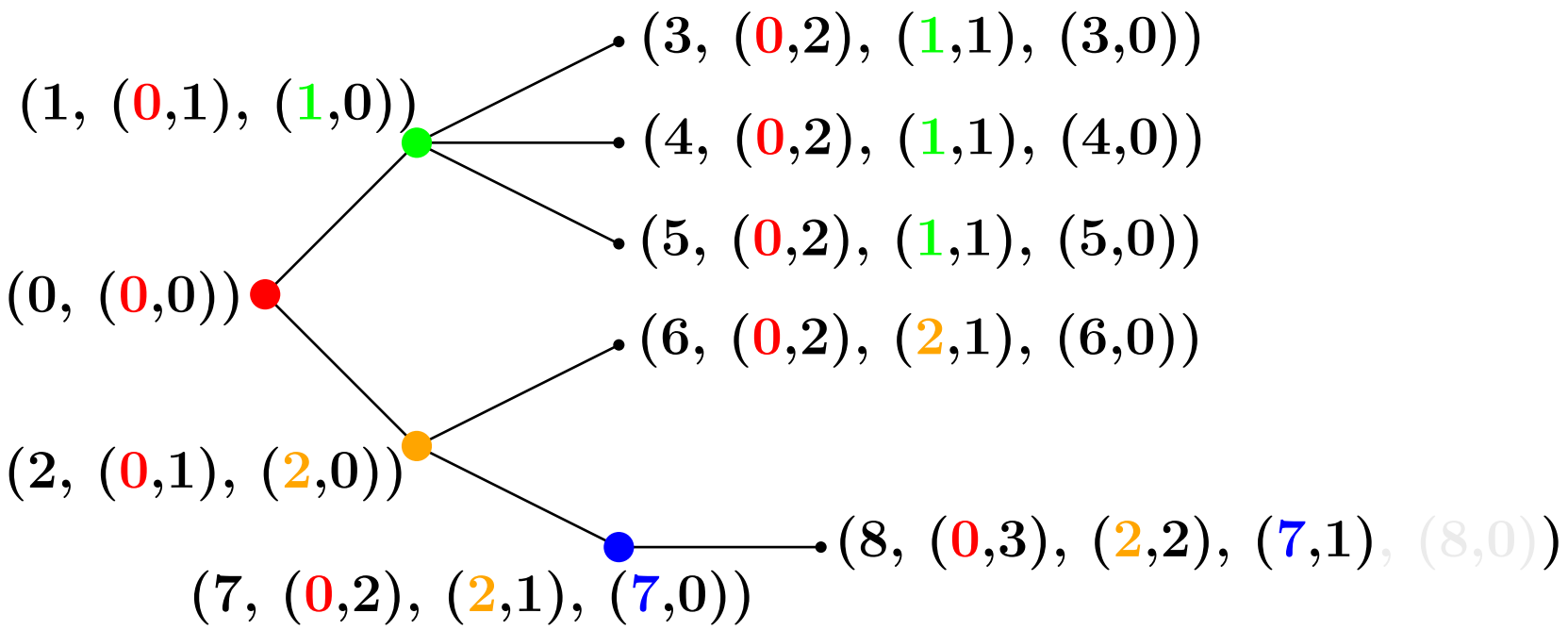
Théorème *(Peleg 2000)*. La famille des arbres sur n sommets admet un DLS avec étiquettes de $O(\log^2 n)$ bits.

DLS pour les arbres : encodage



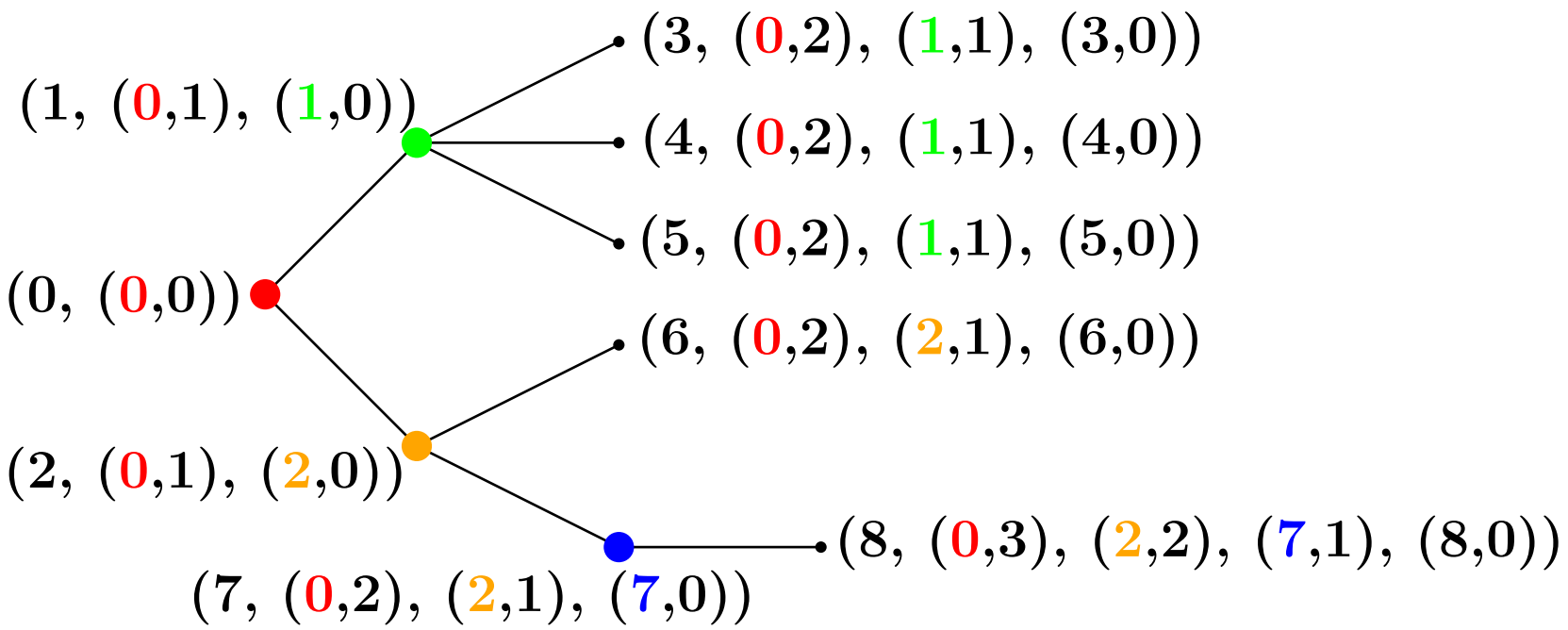
Théorème *(Peleg 2000)*. La famille des arbres sur n sommets admet un DLS avec étiquettes de $O(\log^2 n)$ bits.

DLS pour les arbres : encodage



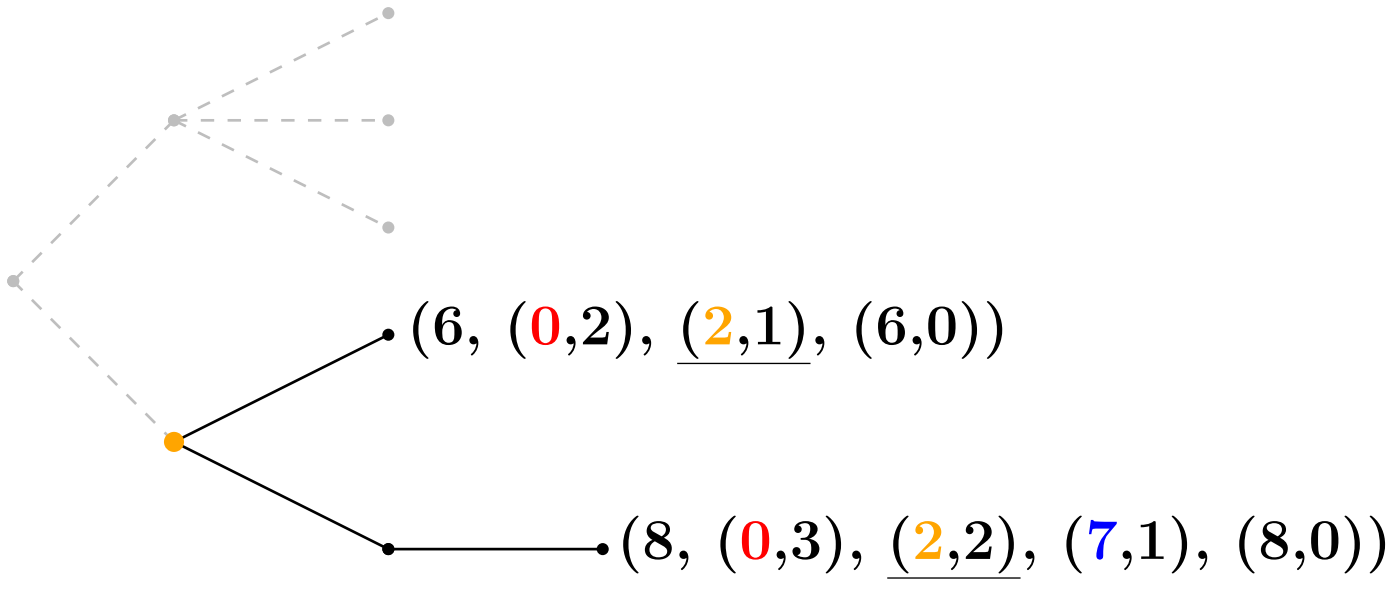
Théorème *(Peleg 2000)*. La famille des arbres sur n sommets admet un DLS avec étiquettes de $O(\log^2 n)$ bits.

DLS pour les arbres : encodage



Théorème *(Peleg 2000)*. La famille des arbres sur n sommets admet un DLS avec étiquettes de $O(\log^2 n)$ bits.

DLS pour les arbres : décodage



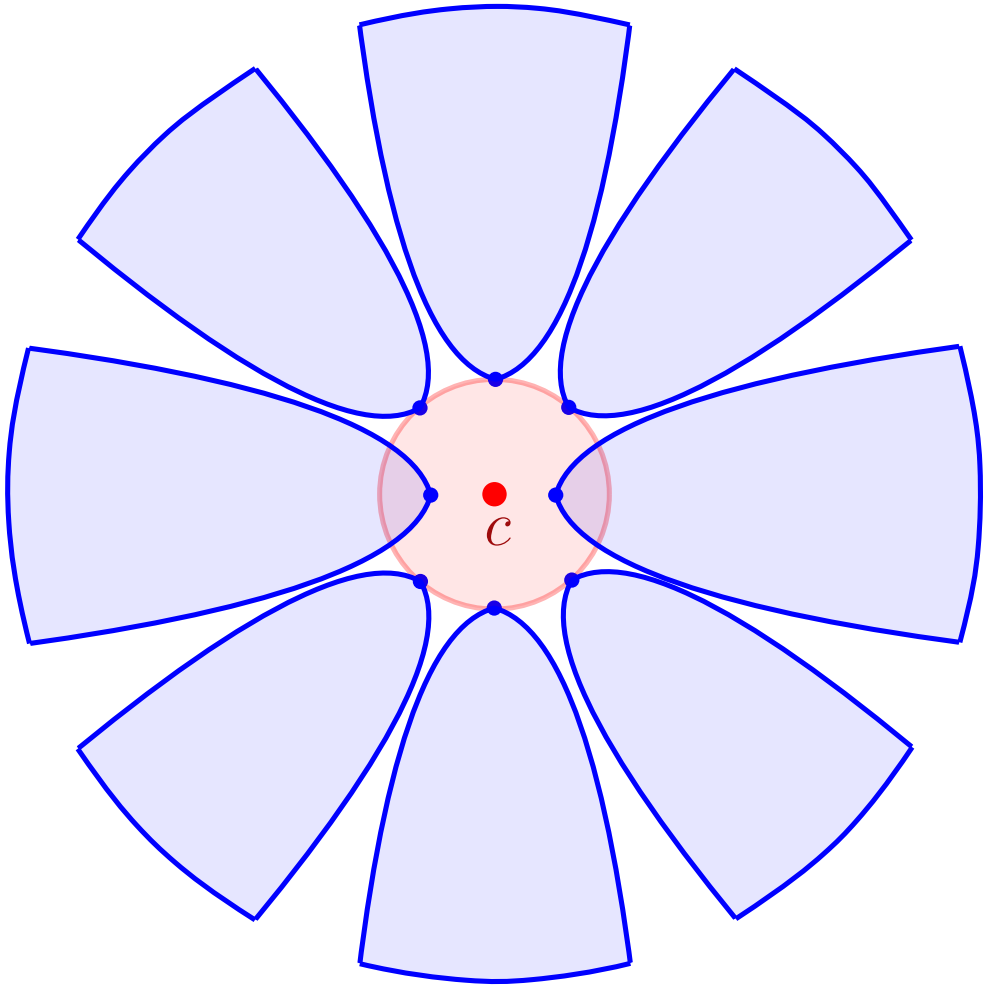
Théorème *(Peleg 2000)*. La famille des arbres sur n sommets admet un DLS avec étiquettes de $O(\log^2 n)$ bits.

Nos résultats

Théorème 3 *(Chepoi, Labourel & R., MFCS 2019)*.
Les graphes médians sans cube ont un DLS avec étiquettes de $O(\log^3 n)$ bits, décodables en temps $O(1)$.

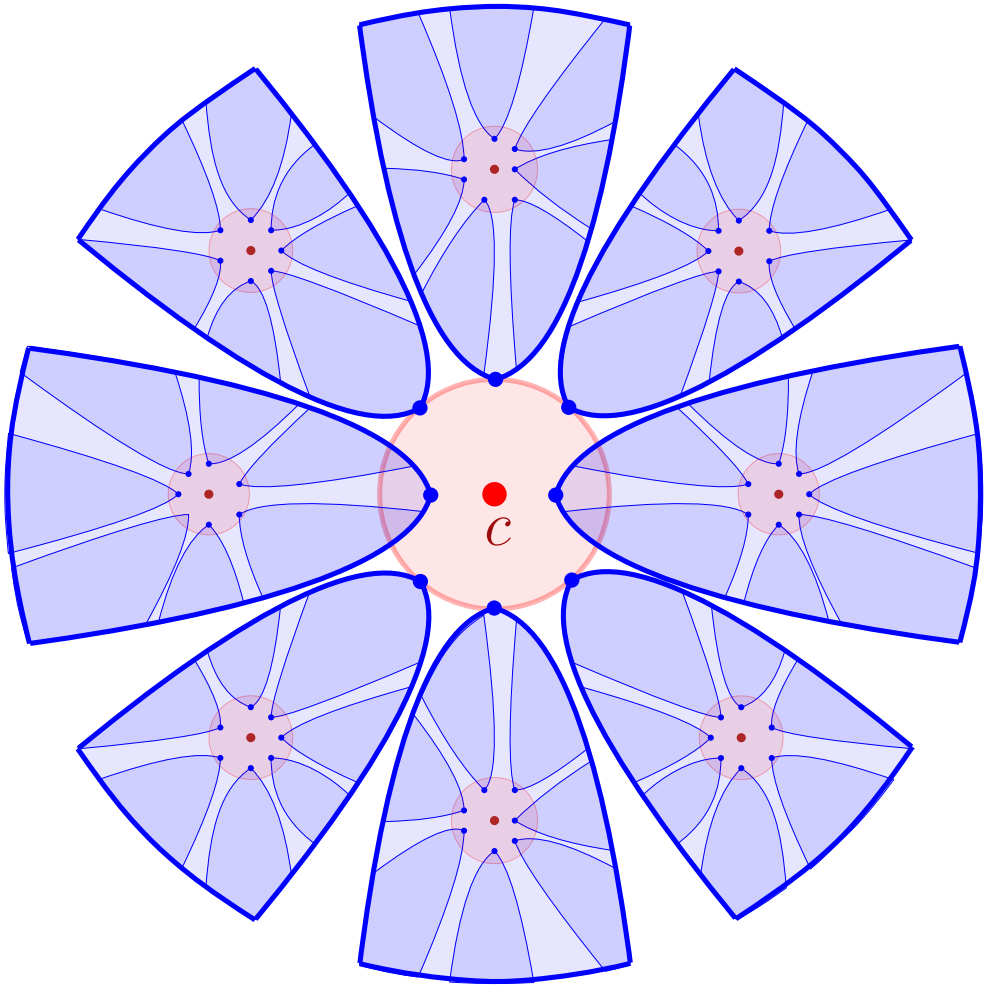
Théorème 4 *(Chepoi, Labourel & R., en préparation)*.
Les graphes pontés sans K_4 ont un DLS 4-approché avec étiquettes de $O(\log^3 n)$ bits, décodables en temps $O(1)$.

Idées principales de nos DLS (et RLS)



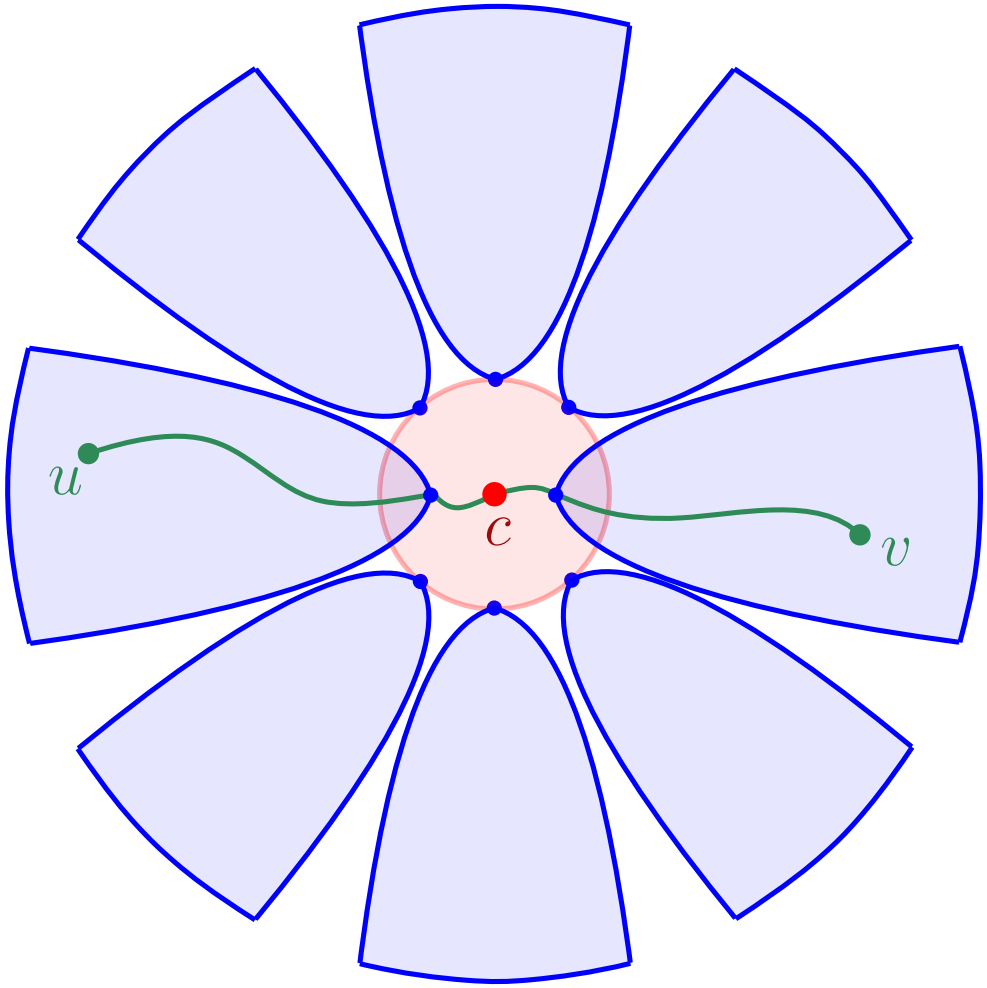
Partitionner G en l'**étoile de c** et les **fibres** de cette étoile.

Idées principales de nos DLS (et RLS)



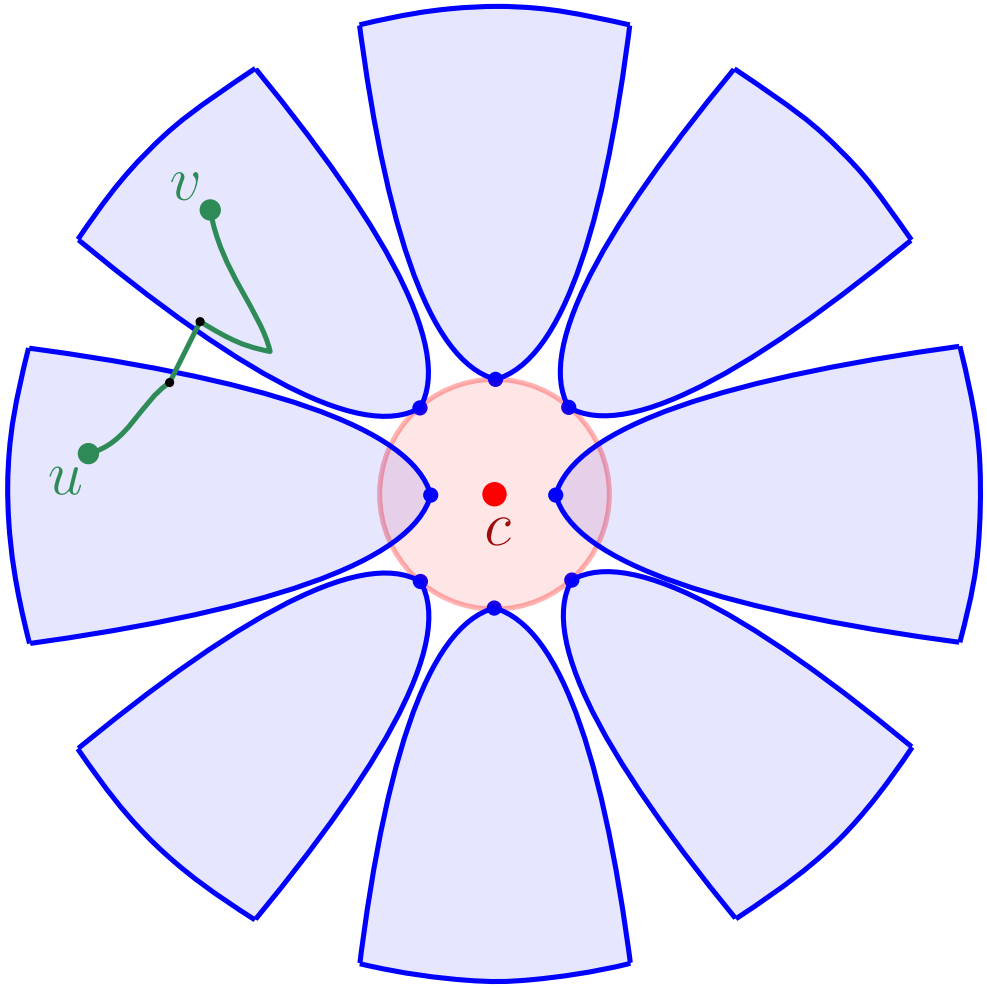
Les fibres restent dans la classe de graphes, sont isométriques et, si c est un **centroïde**, alors elles sont **petites**.

Idées principales de nos DLS (et RLS)



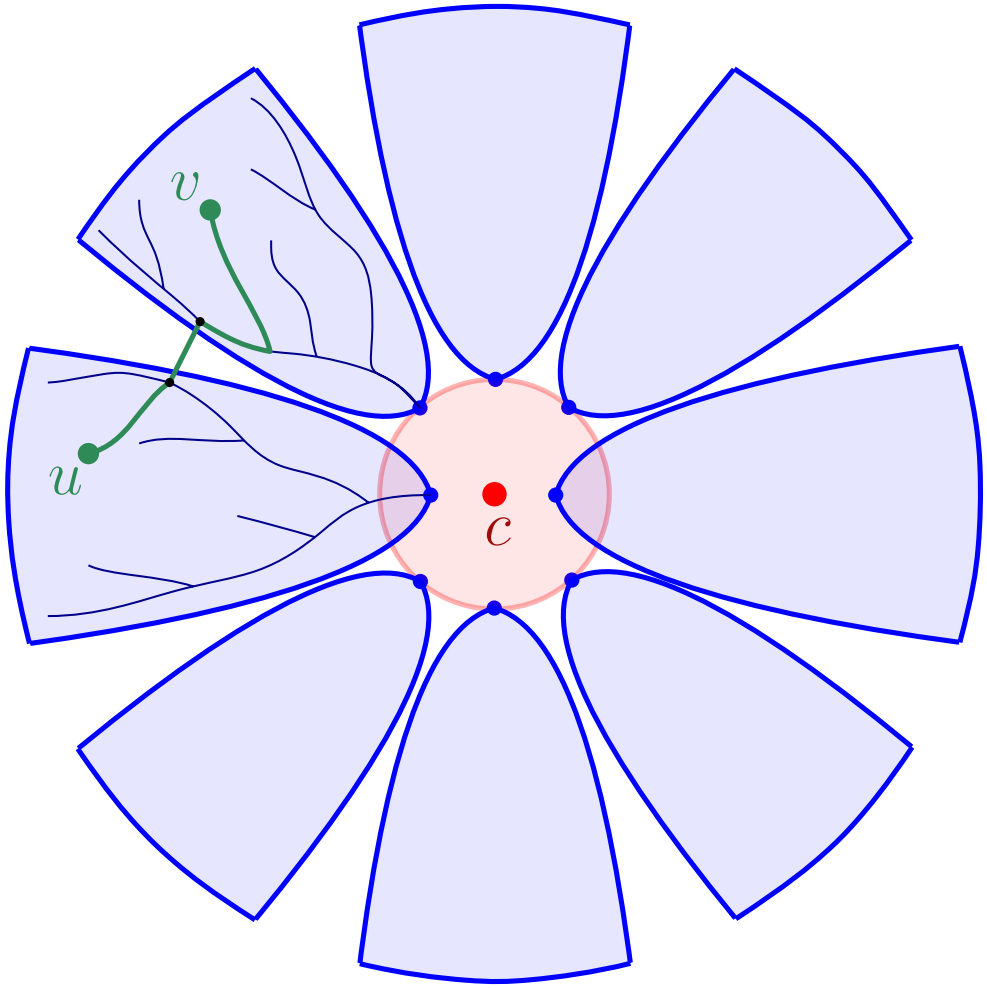
Si les fibres de u et v sont **suffisamment éloignées**, alors c est sur un plus court (u, v)-chemin.

Idées principales de nos DLS (et RLS)



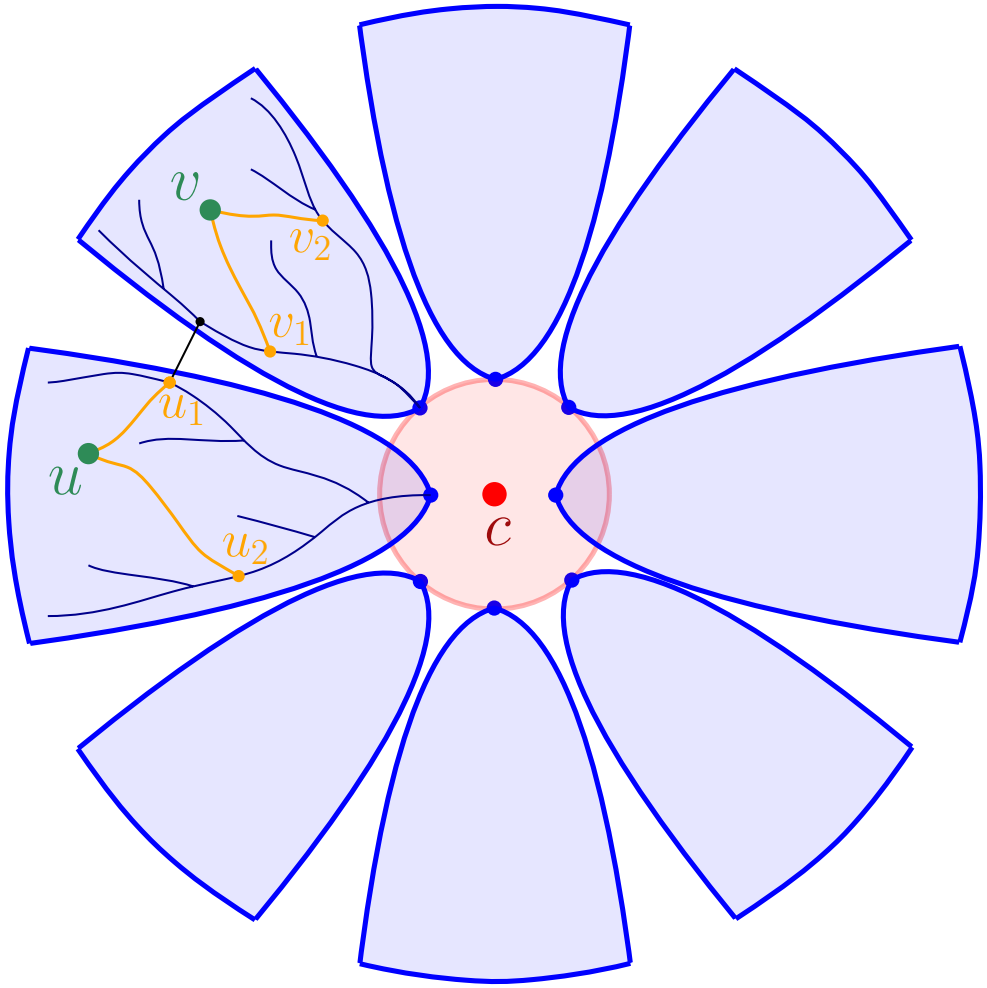
Si les fibres de u et v sont **trop proches**, alors c **n'est pas** sur un plus court (u, v)-chemin.

Idées principales de nos DLS (et RLS)



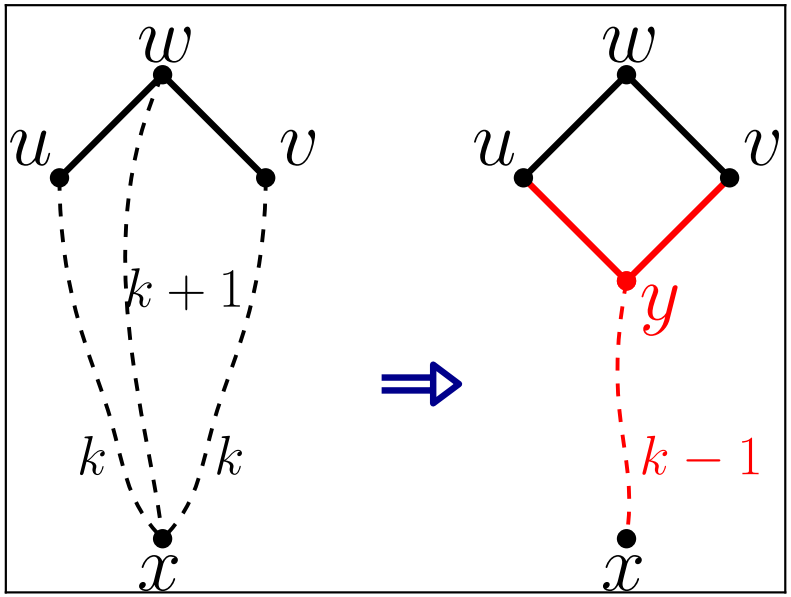
Les **bordures** des fibres sont **presque** des arbres.

Idées principales de nos DLS (et RLS)



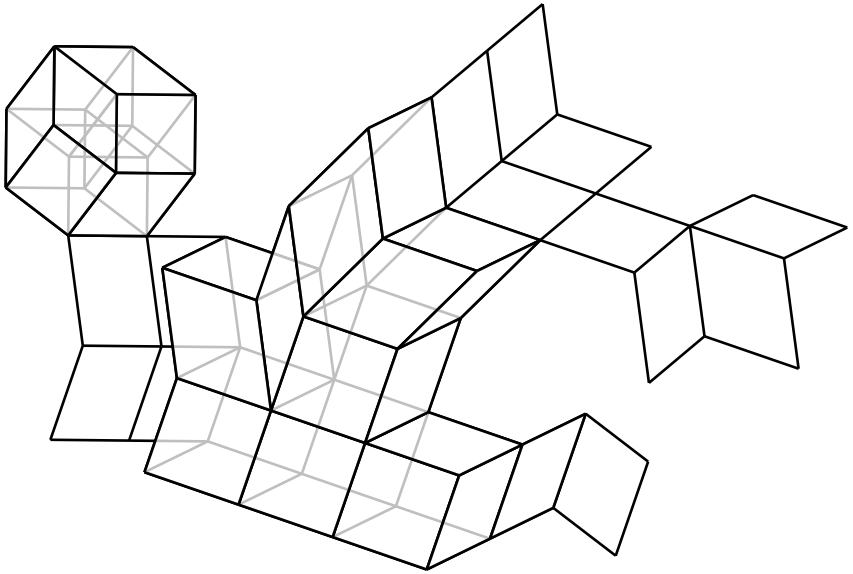
1. Un sommet u possède 2 **sorties** sur la bordure de ses fibres;
2. c ou l'une des sorties est sur un plus court (u, v)-chemin.

Graphes médians



G est **médian** si

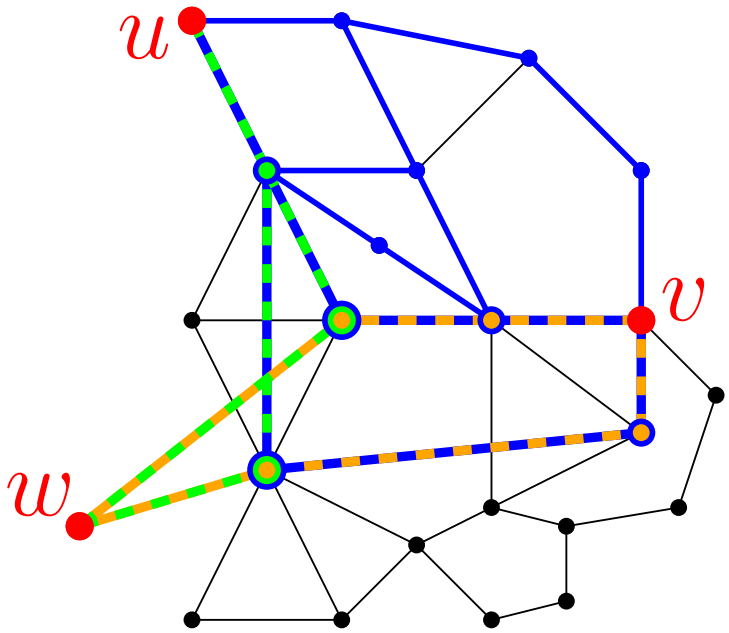
- il est **biparti**;
- il est **sans** $K_{2,3}$;
- il satisfait la **condition du quadrangle**.



Graphes médians 2

Médian de u, v, w : sommet simultanément sur un plus court (u, v)-, (u, w)- et (v, w)-chemin.

G est **médian** si chaque triplet $u, v, w \in V(G)$ admet un **unique** sommet médian.

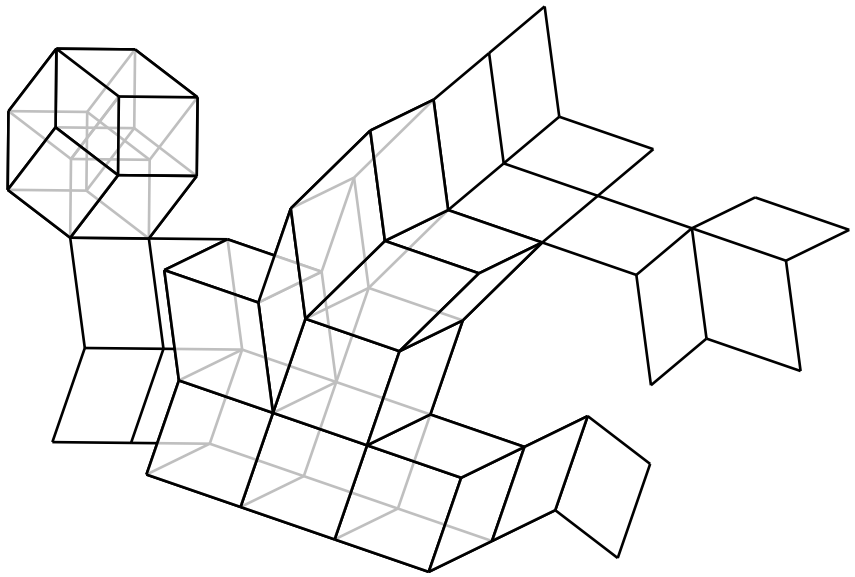


G est **sans cube** s'il est sans Q_3 induit.

Graphes médians 2

Médian de u, v, w : sommet simultanément sur un plus court (u, v)-, (u, w)- et (v, w)-chemin.

G est **médian** si chaque triplet $u, v, w \in V(G)$ admet un **unique** sommet médian.

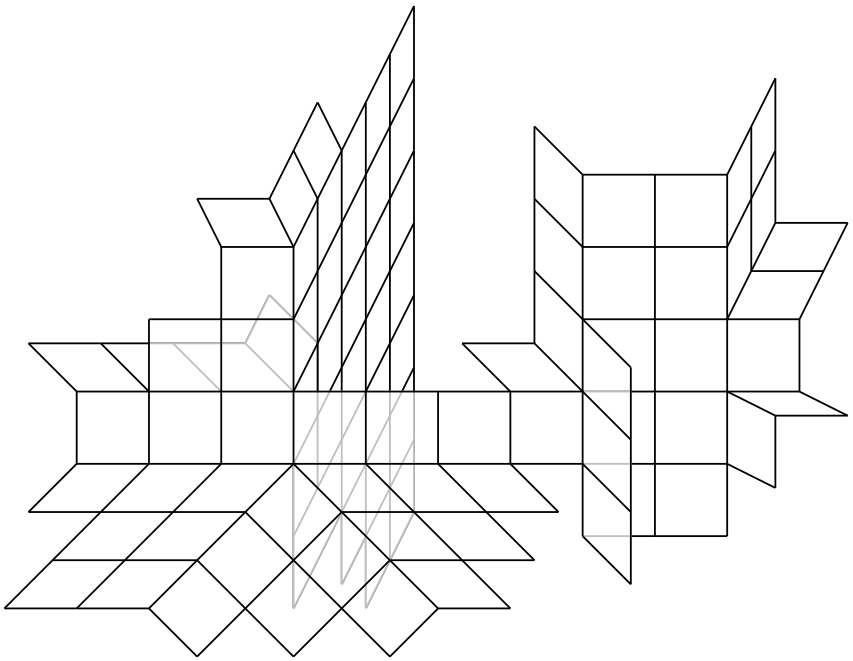


G est **sans cube** s'il est sans Q_3 induit.

Graphes médians 2

Médian de u, v, w : sommet simultanément sur un plus court (u, v)-, (u, w)- et (v, w)-chemin.

G est **médian** si chaque triplet $u, v, w \in V(G)$ admet un **unique** sommet médian.

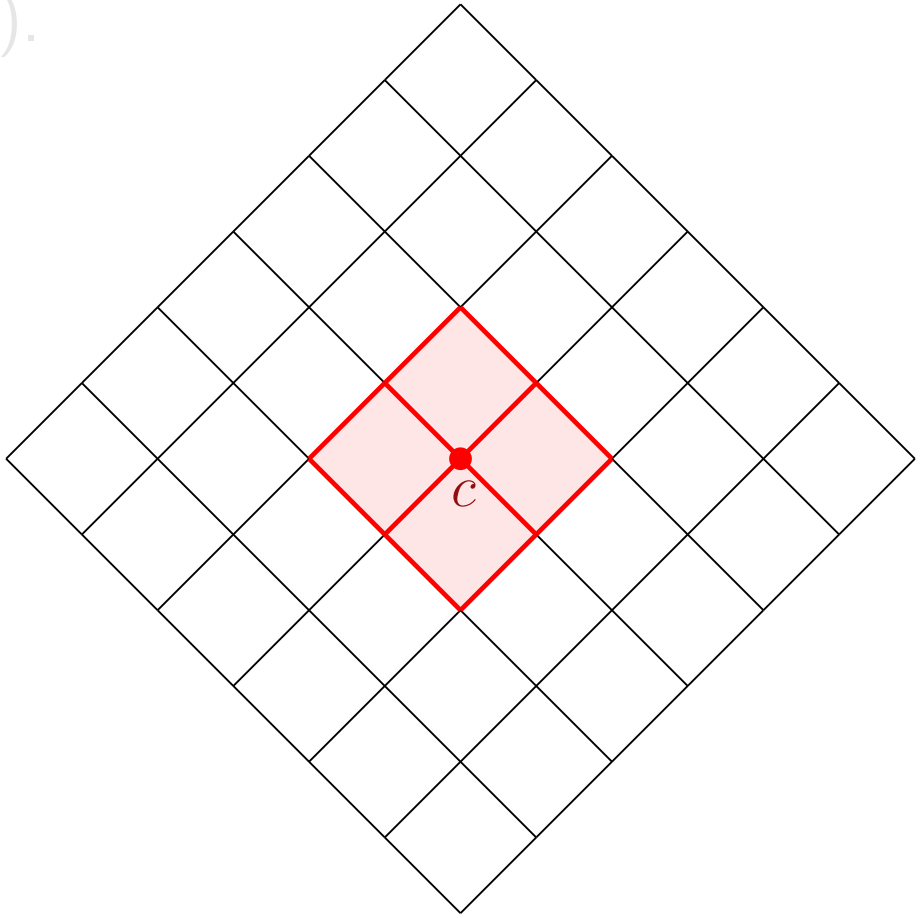


G est **sans cube** s'il est sans Q_3 induit.

Partition de graphes médians

Étoile de $\mathrm{St}(c)$: arêtes et carrés incidents à c.

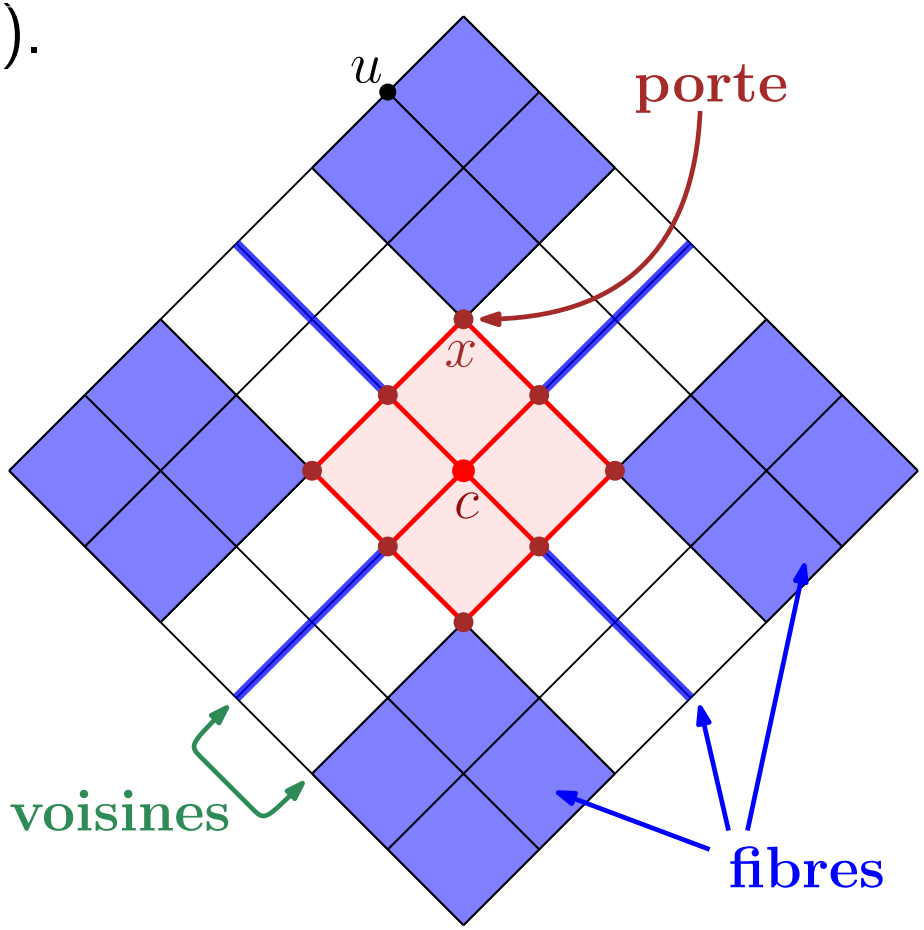
Fibre $F(x)$ de $x \in \mathrm{St}(c)$: tous les sommets ayant x pour porte sur $\mathrm{St}(c)$.



Partition de graphes médians

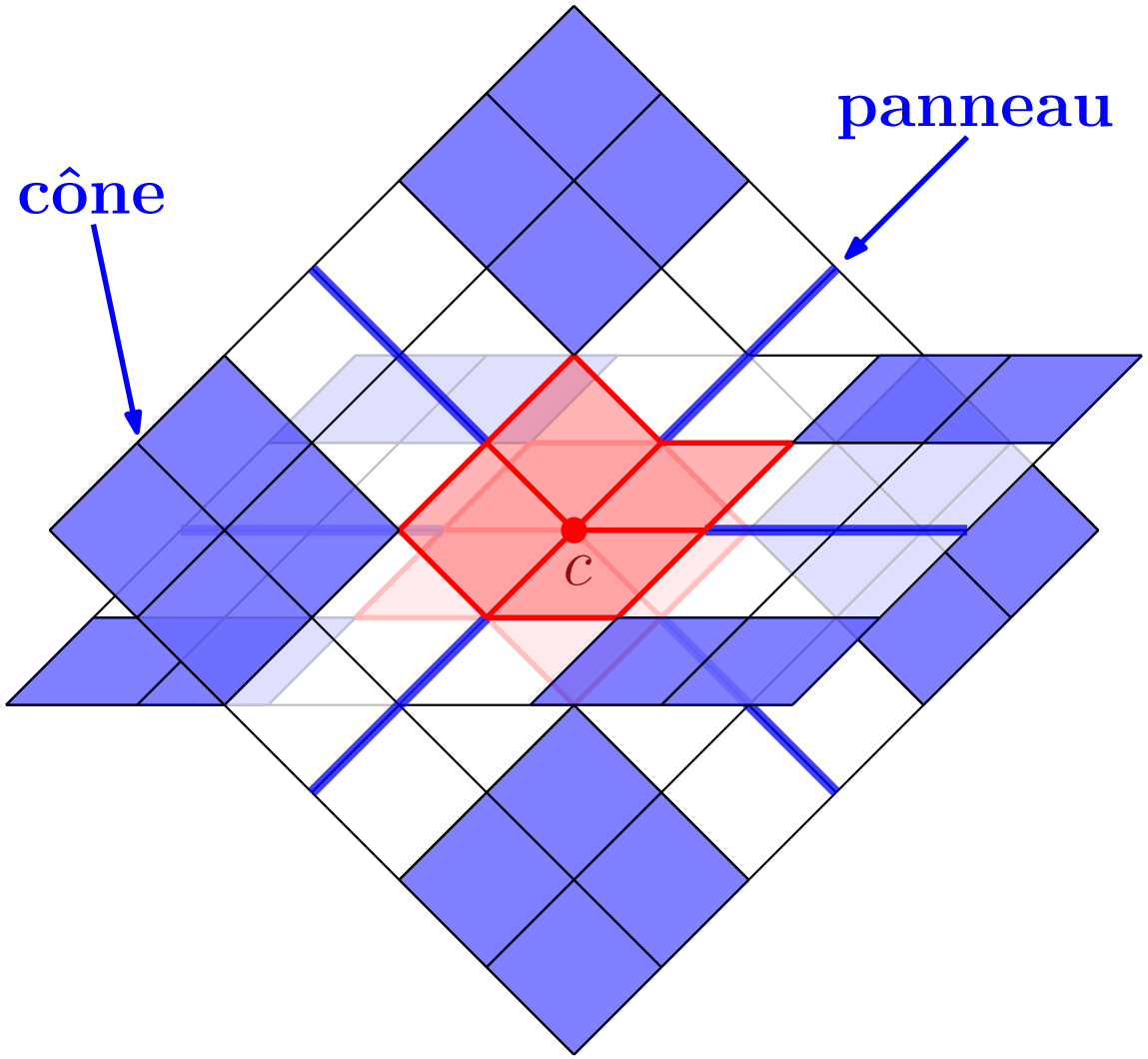
Étoile de $\mathrm{St}(c)$: arêtes et carrés incidents à c.

Fibre $F(x)$ de $x \in \mathrm{St}(c)$: tous les sommets ayant x pour porte sur $\mathrm{St}(c)$.



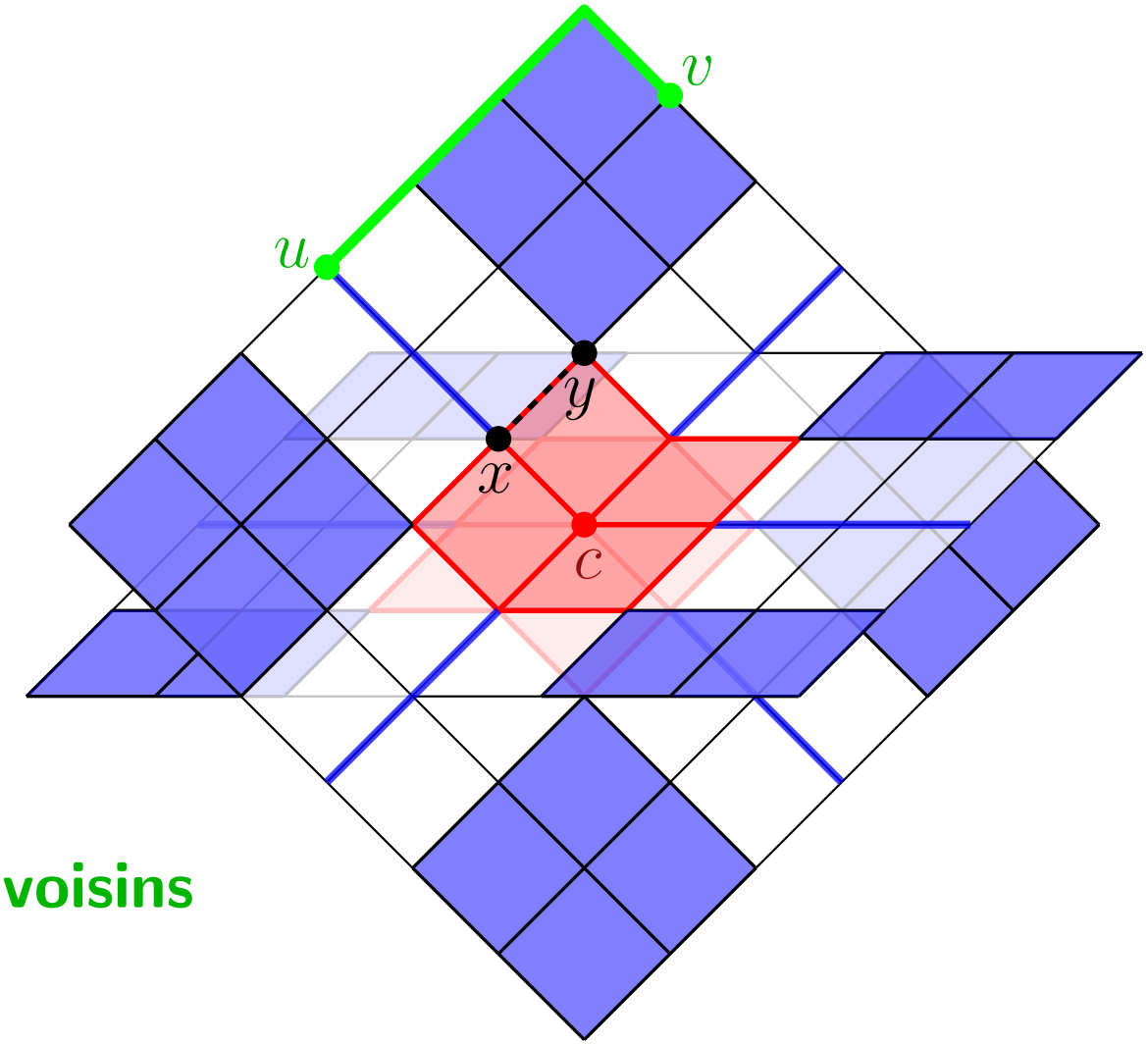
Types des fibres

Graphes médians sans cube



Plus courts chemins

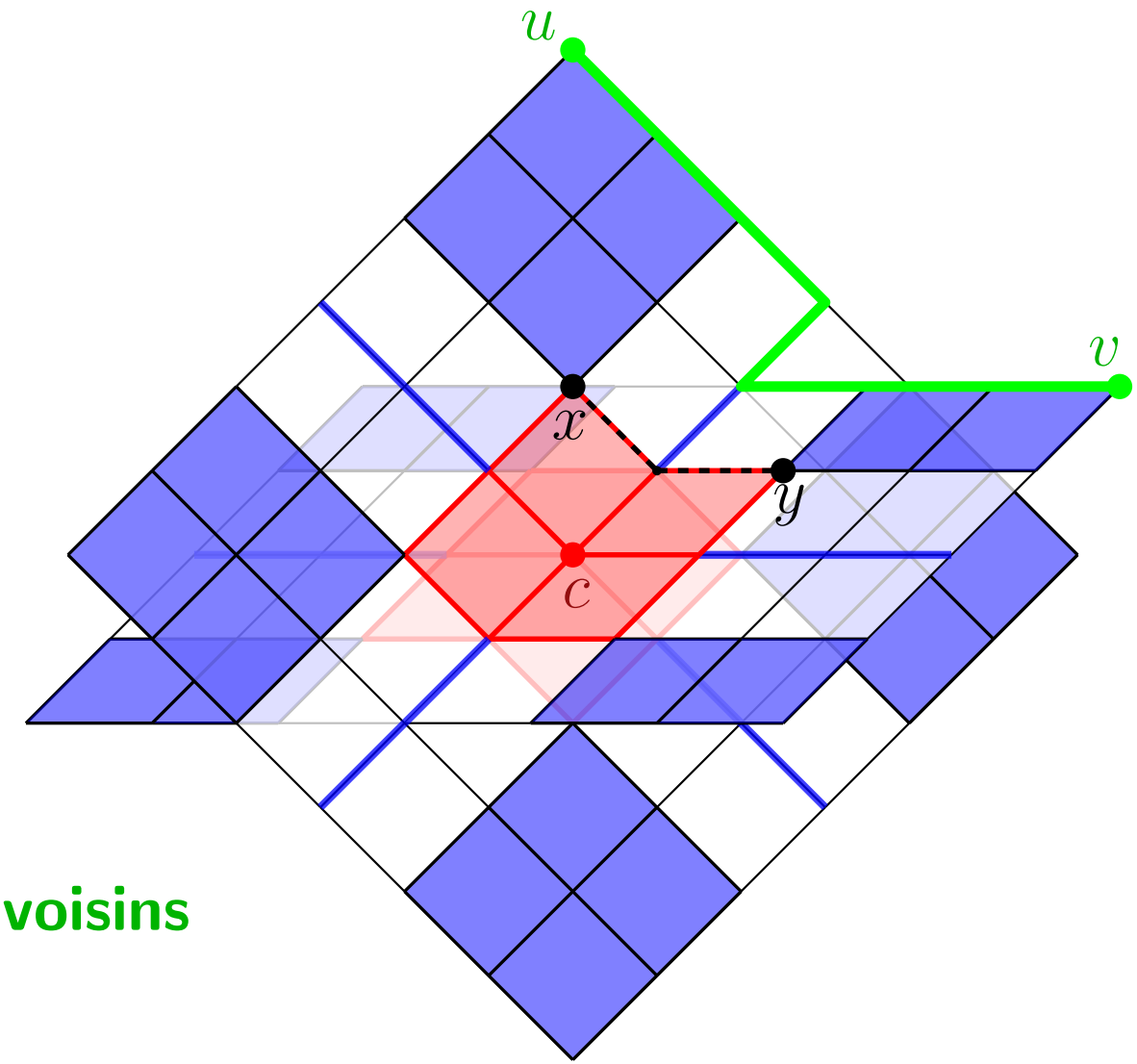
Graphes médians sans cube



Sommets **1-voisins**

Plus courts chemins

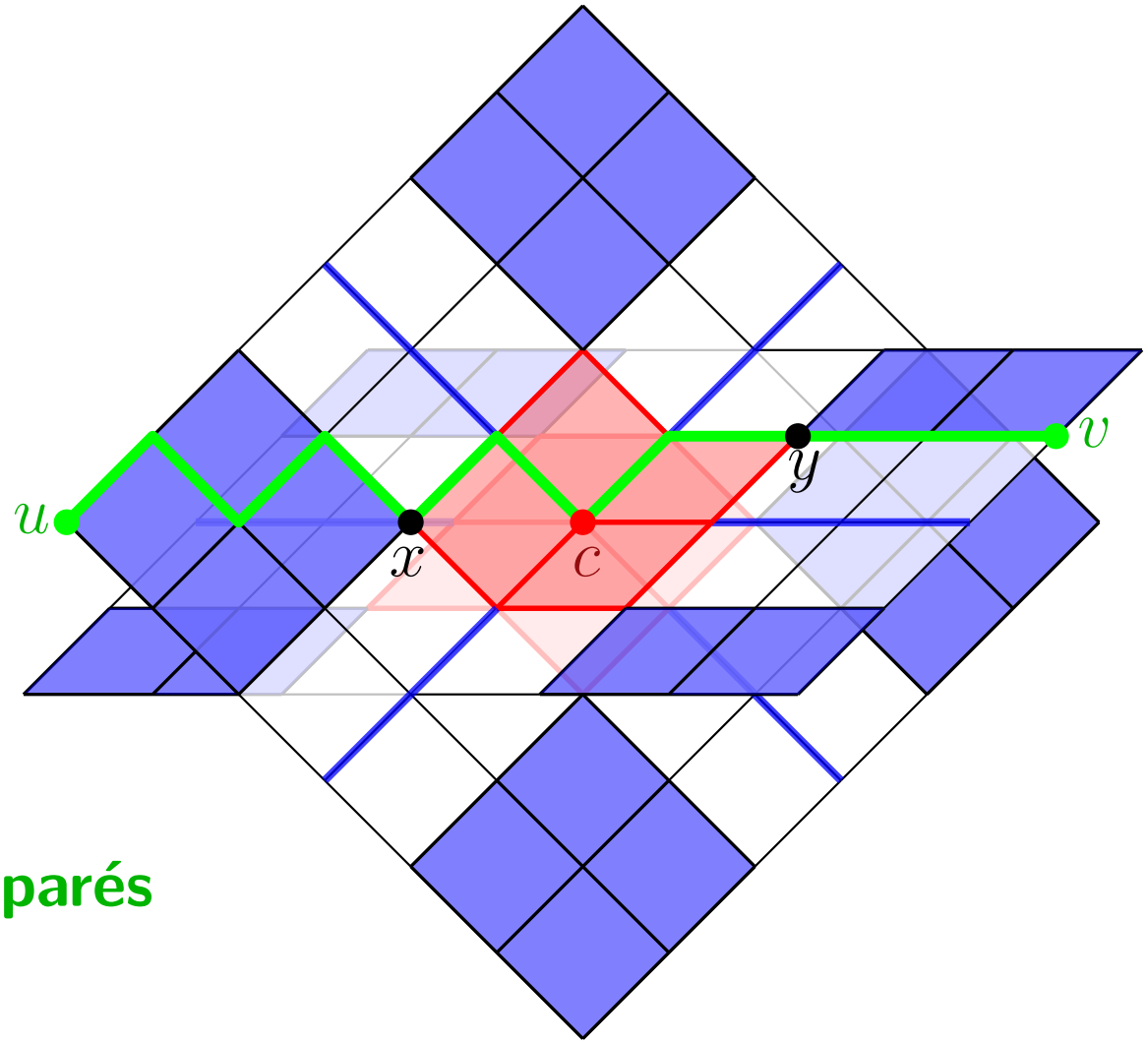
Graphes médians sans cube



Sommets **2-voisins**

Plus courts chemins

Graphes médians sans cube



Sommets **séparés**

Encodage des graphes médians sans cube

Théorème 3 : Étiquettes de $O(\log^3 n)$ bits.

ENCODAGE.

(1) Attribuer à chaque sommet u de G un identifiant $\mathrm{id}(u)$;

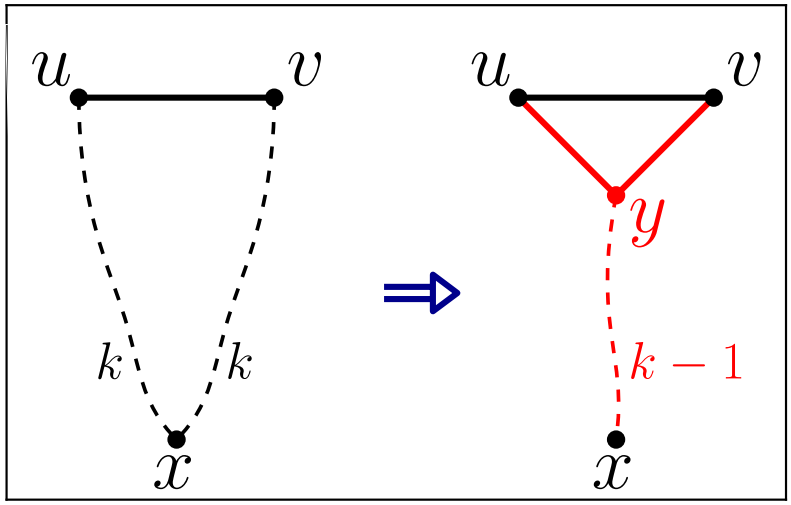
(2) Trouver un centroïde c de G et construire la partition;

(3) Concaténer les informations suivantes à l'étiquette de chaque sommet u:

- $\mathrm{id}(c)$, $\mathrm{d}(u, c)$, $\mathrm{L}_{\mathrm{St}(c)}(x)$; *$x$ porte de u sur $\mathrm{St}(c)$*
- $\mathrm{d}(u, u_1)$, $\mathrm{L}_{T_1}(u_1)$; *T_1, T_2 sont 2 bordures*
- $\mathrm{d}(u, u_2)$, $\mathrm{L}_{T_2}(u_2)$; *u_1, u_2 sont les sorties de u*

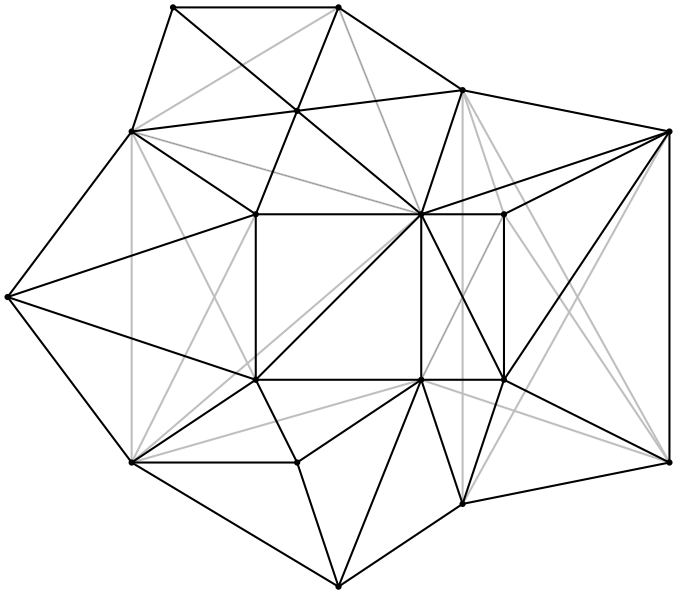
(4) Retourner en (2) pour chaque fibre non triviale de la partition.

Graphes pontés

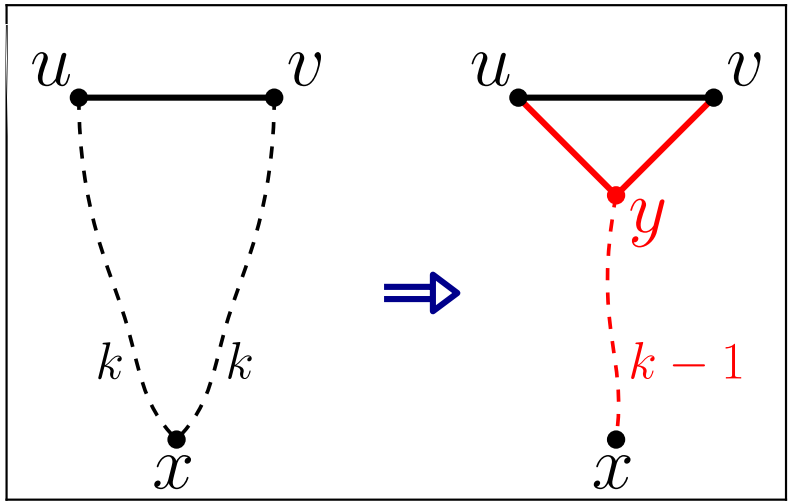


G est **ponté** si

- il est **sans** C_4;
- il est **sans** C_5;
- il satisfait la **condition du triangle**;
- il satisfait la **condition du quadrangle**.

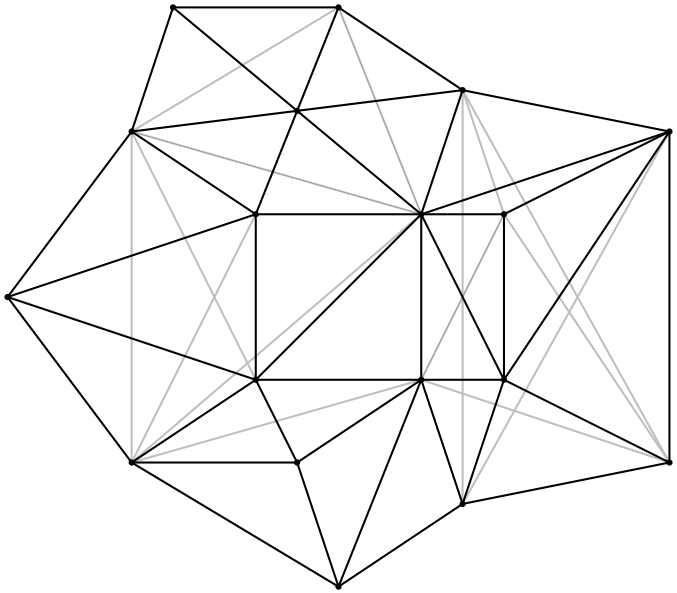


Graphes pontés



G est **ponté** si

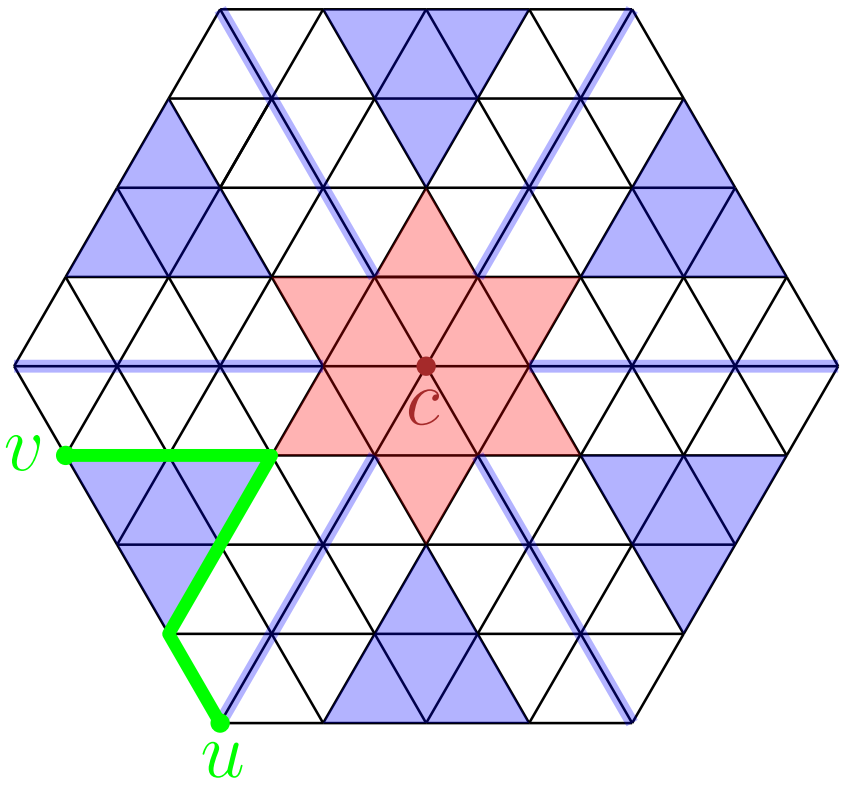
- il est **sans** C_4;
- il est **sans** C_5;
- il satisfait la **condition du triangle**;
- il satisfait la **condition du quadrangle**.



Également.
G est **ponté** ssi il ne contient pas de cycle isométrique de longueur > 3.

Plus courts chemins approchés

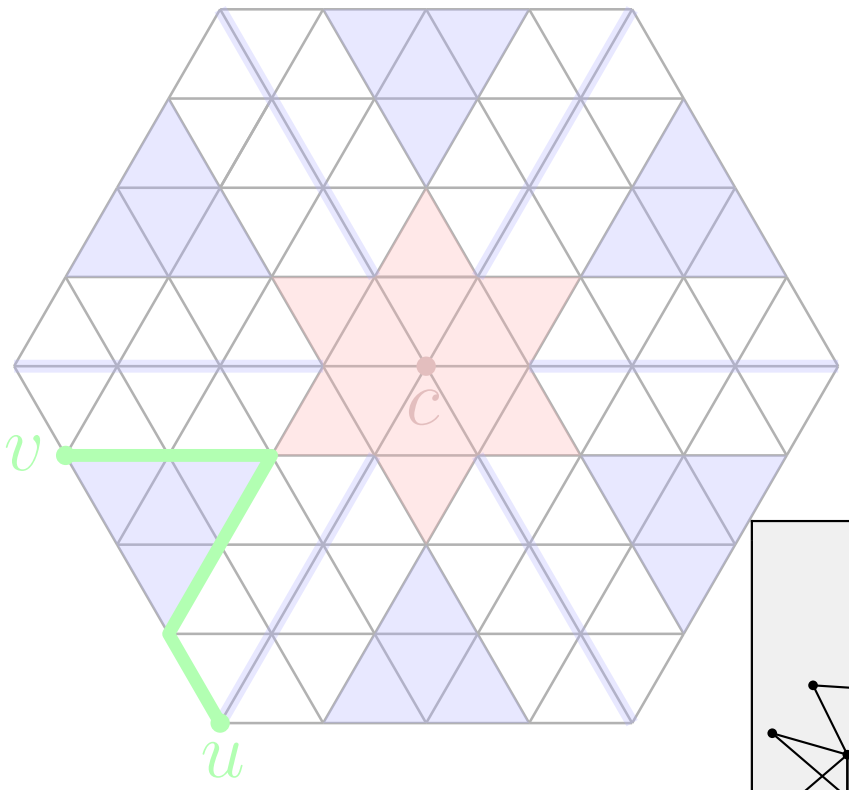
Graphes pontés sans K_4

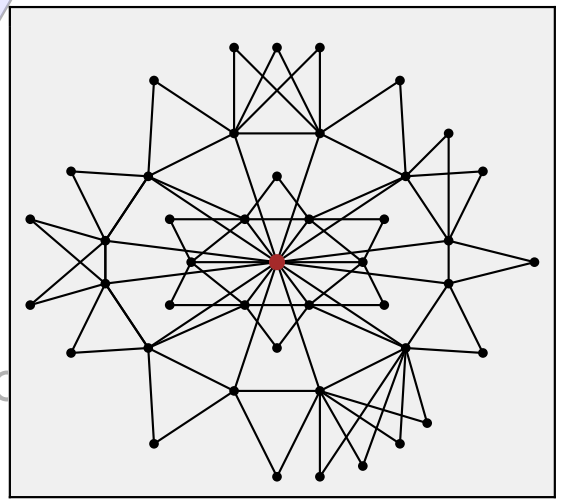


rappel: DLS 4-approché pour les graphes pontés sans K_4.

Plus courts chemins approchés

Graphes pontés sans K_4





rappel: DLS 4-approché pour les graphes po

Conclusion et Questions (1/2)

Résultats.

1. Si $G = (V, E)$ est un sous-graphe de produit cartésien de graphes sans mineur H, alors $\frac{|E|}{|V|} \leq c_H \cdot \text{VC-dim}^*(G)$.
2. Si $G = (\mathcal{S}, E)$ est un sous-graphe de demi-cube, alors $\frac{|E|}{|\mathcal{S}|} \leq \binom{\text{cVC-dim}^*(\mathcal{S})}{2}$.

Questions.

1. Bornes plus serrées sur la densité de ces graphes ?
   - G sous-graphe de produit cartésien : $\frac{|E|}{|V|} \leq \text{VC-dens}^*(G)$?
   - G sous-graphe de demi-cube : $\frac{|E|}{|\mathcal{S}|} = O(\omega(G) \cdot \text{VC-dim}(\mathcal{S}))$?
2. Quelles sont les portées de nos notions de VC-dimension ?

Conclusion et Questions (2/2)

Résultats.

1. DLS pour les graphes médians sans cubes.
2. DLS 4-approchés pour les graphes pontés sans K_4.

Questions.

1. Analyse plus fine du facteur d'approximation de notre DLS pour les graphes pontés sans K_4 ?
2. DLS (et RLS) avec étiquettes poly-logarithmiques pour les graphes **médians généraux** ? **hyperboliques** ?
3. DLS (et RLS) exacts ou approchés avec étiquettes poly-logarithmiques pour les graphes **pontés généraux** ?