Recherche de motifs dans les graphes colorés

Florian Sikora

LAMSADE

8 octobre 2012

Plan

Introduction

Complexité paramétrée

Recherche exacte de motifs

Recherche approchée de motifs

Perspectives

Plan

Introduction

Complexité paramétrée

Recherche exacte de motifs

Recherche approchée de motifs

Perspectives

Motivations

25000 30000 45000

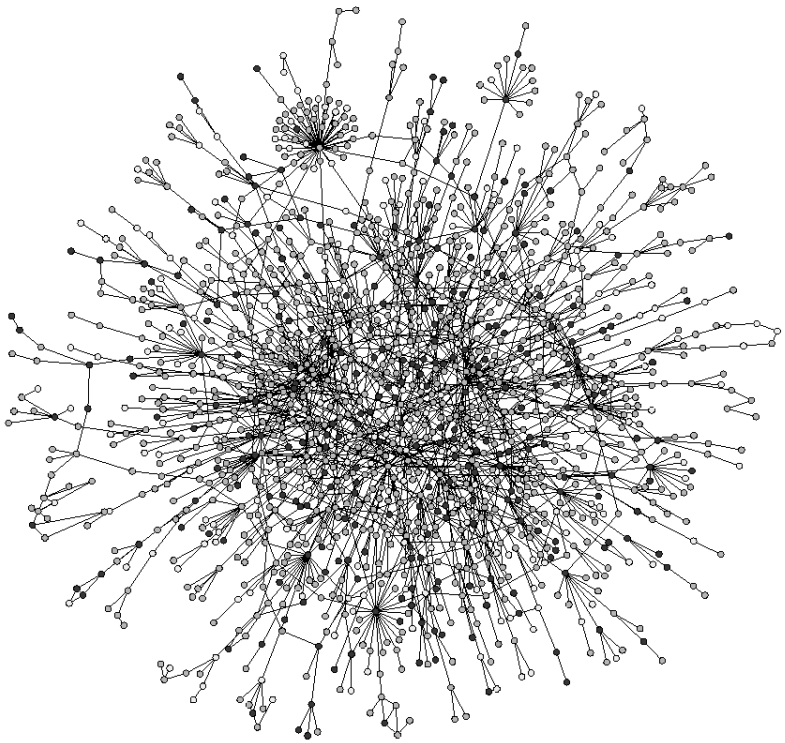
- Complexité de l'homme $\not\Leftrightarrow$ nombre de gènes ?

Réseaux biologiques

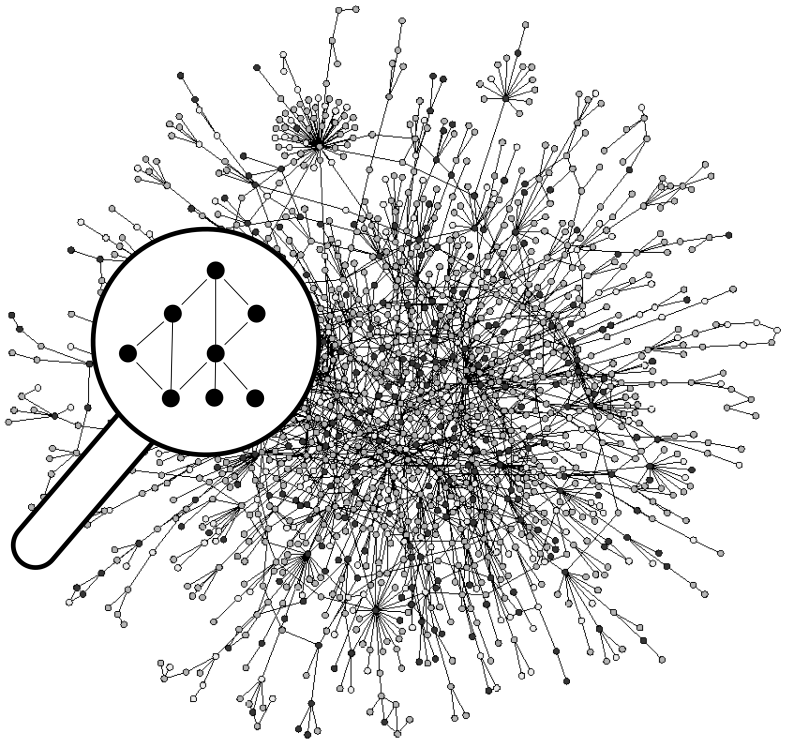
- Énumérer les composants d'un avion ne suffit pas pour comprendre son fonctionnement.
- Étudier les relations entre les éléments.

Réseaux biologiques

- Les éléments biologiques (protéines, métabolites...) peuvent interagir entre-eux.



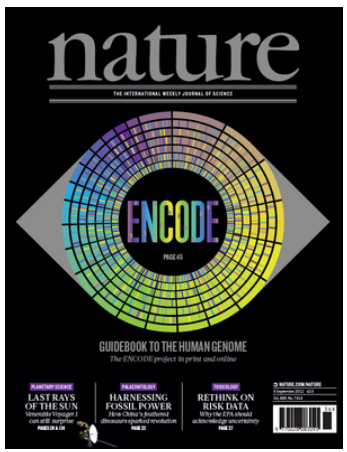
Réseaux biologiques



- Modélisation par un **graphe**.
  - Les **éléments** sont représentés par les **nœuds**.
  - Les **interactions** sont représentées par les **arêtes**.

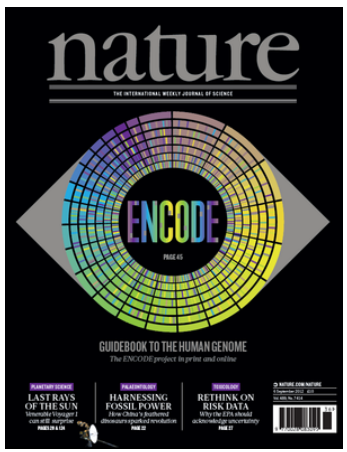
Enjeux

- Nouvelles techniques : le nombre de données augmente très rapidement [Sharan et Ideker 2006].
  - 2001 : quelques centaines d'interactions.
  - 2006 : plusieurs milliers.
- Beaucoup de bases de données :
  - BIND,
  - DIP,
  - KEGG,
  - MINT,
  - ...
- Encyclopédie ENCODE [Nature Sept 2012].



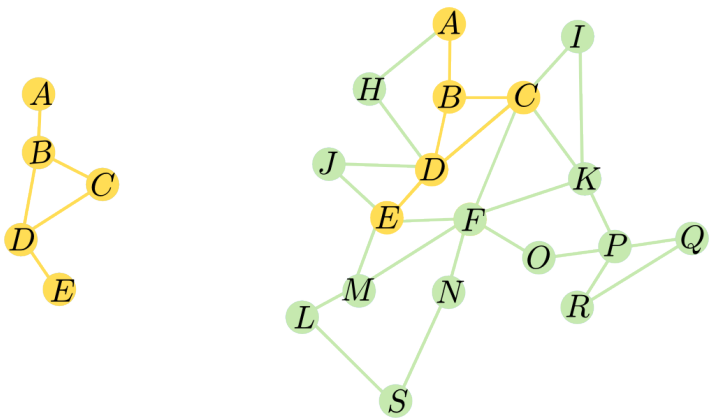
Enjeux

- Nouvelles techniques : le nombre de données augmente très rapidement [Sharan et Ideker 2006].
  - 2001 : quelques centaines d'interactions.
  - 2006 : plusieurs milliers.
- Beaucoup de bases de données :
  - BIND,
  - DIP,
  - KEGG,
  - MINT,
  - ...
- Encyclopédie ENCODE [Nature Sept 2012].
- Besoin de l'outil informatique.



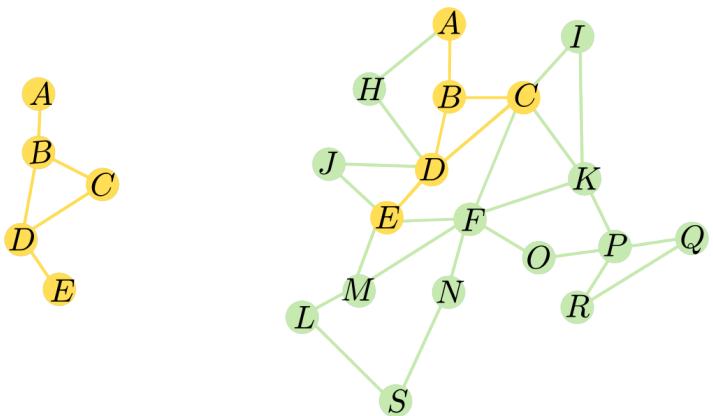
Rechercher des motifs

- Retrouver un sous-réseau respectant les labels et la topologie du motif.
- Deux labels correspondent si les éléments ont une fonction similaire (par ex. homologies pour les protéines).



Rechercher des motifs

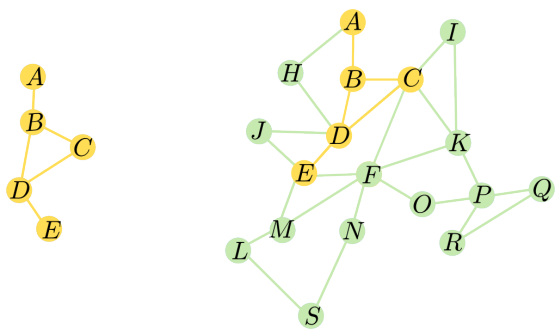
- Retrouver un sous-réseau respectant les labels et la topologie du motif.
- Deux labels correspondent si les éléments ont une fonction similaire (par ex. homologies pour les protéines).



- Chercher des motifs pour retrouver des fonctions connues.
- **Déduire** les informations d'espèces **peu connues** depuis des espèces **bien connues**.

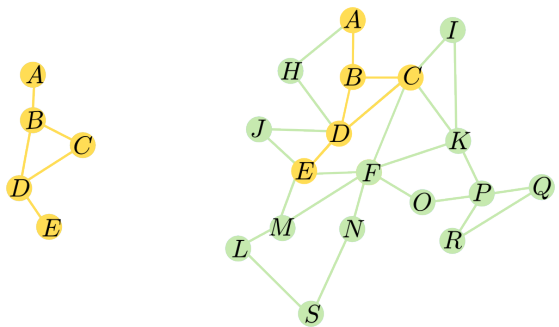
Rechercher des motifs

- Problèmes rapidement **NP-complets**.
  - N'existe probablement **pas d'algorithme exact** de complexité pire cas **polynomiale** (par ex. $\mathcal{O}(n^c)$).
  - Complexité **exponentielle** (par ex. $\mathcal{O}(c^n)$) pour au moins une instance.



Rechercher des motifs

- Mais tout n'est pas perdu !
- Idée : exploiter le fait que les **motifs sont plus petits** ($\sim$ 5 à 15 nœuds) que le réseau (par ex. $\sim$ 5.000 pour la levure).
- Confiner la partie exponentielle à k (taille motif) au lieu de n (taille du réseau) : sujet de la **complexité paramétrée**.



Plan

Introduction

Complexité paramétrée

Recherche exacte de motifs

Recherche approchée de motifs

Perspectives

Algorithmes FPT

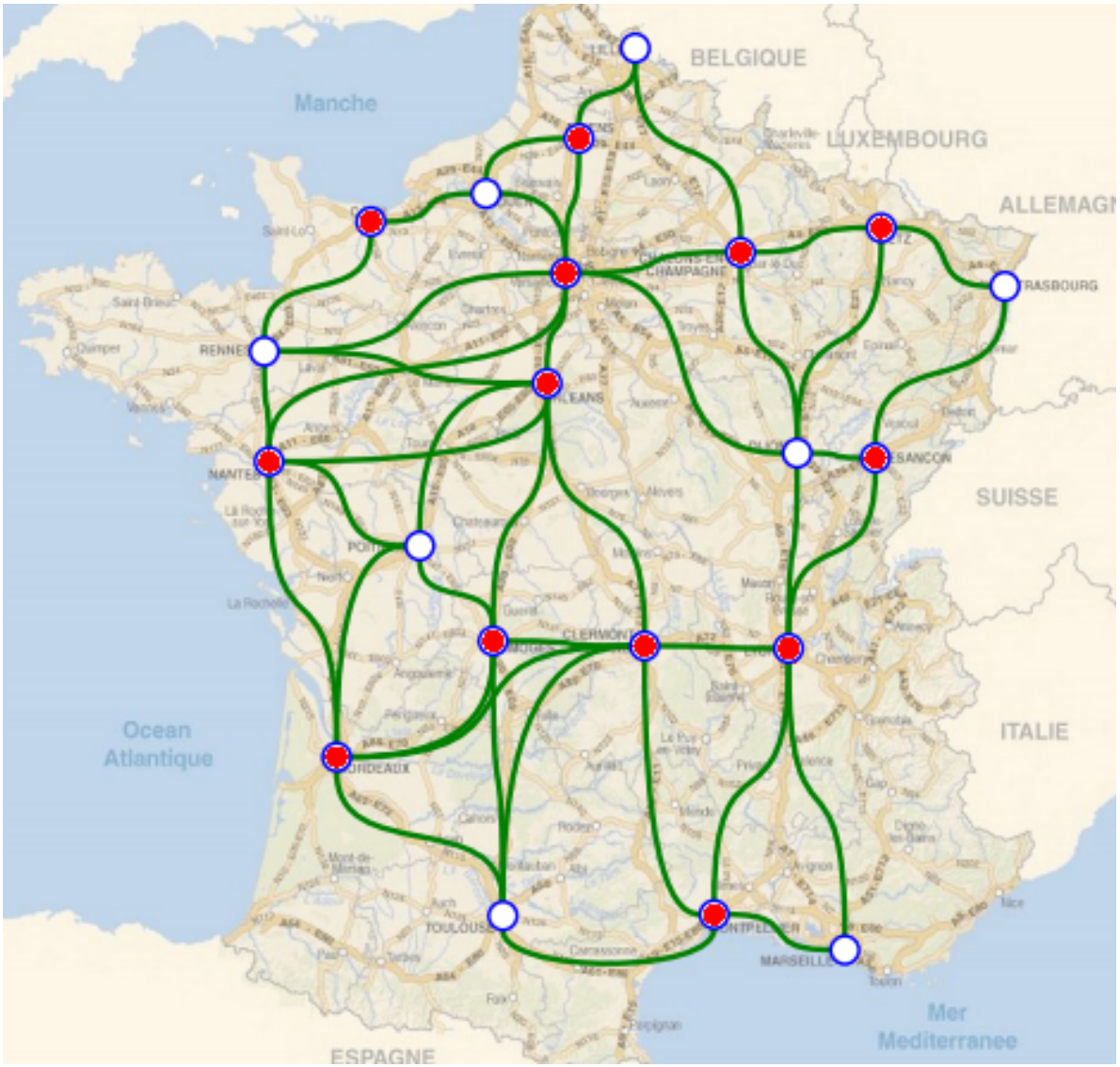
- Beaucoup de problèmes (paramétrés) sont de la forme :
  - Entrée : Un objet X, $|X| = n$, un entier k.
  - Question : Est-ce que X a une propriété selon k ?

Algorithmes FPT

- Beaucoup de problèmes (paramétrés) sont de la forme :
  - Entrée : Un objet X, $|X| = n$, un entier k.
  - Question : Est-ce que X a une propriété selon k ?
- Plusieurs paramètres possibles (souvent la taille de la solution).

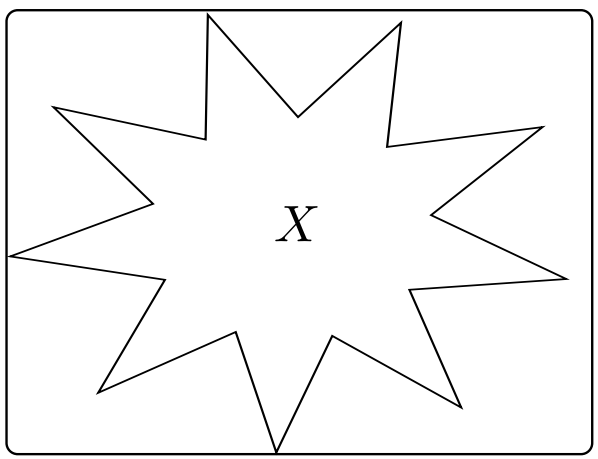
Exemple : Vertex Cover

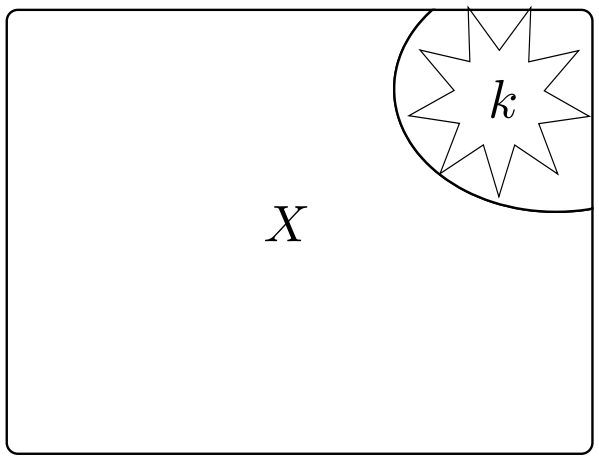
Exemple : Vertex Cover



Complexité paramétrée

- Problème dans la classe FPT s'il existe un algorithme **exact exponentiel** seulement en son **paramètre k** (et pas en la taille de l'entrée n) [Downey et Fellows 1999]. Exemple Vertex Cover en $\mathcal{O}^*(1.28^k)$.





Algorithme exact avec complexité pire cas de la forme $f(k) \cdot |X|^c$

Complexité paramétrée

- Problème dans la classe FPT s'il existe un algorithme **exact exponentiel** seulement en son **paramètre k** (et pas en la taille de l'entrée n) [Downey et Fellows 1999]. Exemple Vertex Cover en $\mathcal{O}^*(1.28^k)$.
- Parfois impossible : W-difficulté. Exemple Independent Set.

Approximation

- Un autre possibilité pour contourner la difficulté d'un problème : l'approximation.

Approximation

- Un autre possibilité pour contourner la difficulté d'un problème : l'approximation.
- Pour un problème d'optimisation NP-difficile, pas d'algorithme **polynomial** possible (sauf si P = NP)...
- ... si la solution doit être optimale !

Approximation

- Un autre possibilité pour contourner la difficulté d'un problème : l'approximation.
- Pour un problème d'optimisation NP-difficile, pas d'algorithme **polynomial** possible (sauf si P = NP)...
- ... si la solution doit être optimale !
- Sacrifier de l'exactitude pour obtenir un algorithme polynomial.

Plan

Introduction

Complexité paramétrée

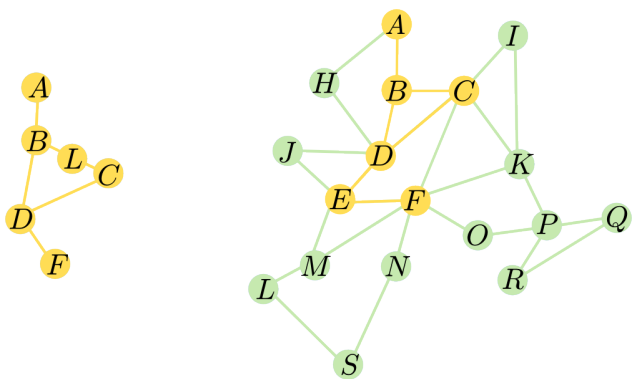
Recherche exacte de motifs

Recherche approchée de motifs

Perspectives

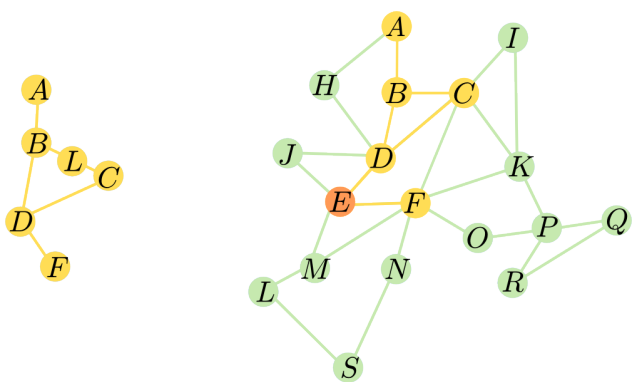
Recherche de motifs

- Vision "classique" de la recherche de motifs :
  - Le motif a une certaine topologie (chemin, arbre, graphes spécifiques...).
- La solution correspond aux labels **et** à la topologie du motif.



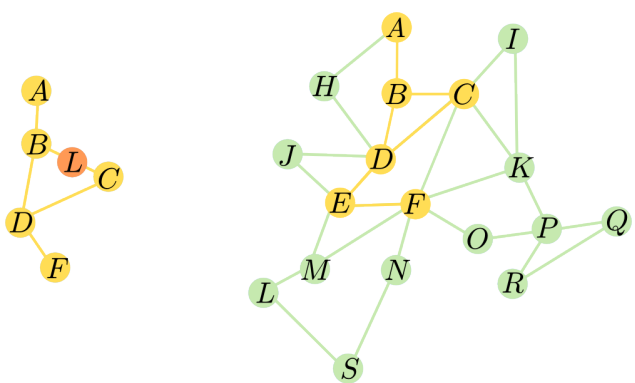
Recherche de motifs

- Vision "classique" de la recherche de motifs :
  - Le motif a une certaine topologie (chemin, arbre, graphes spécifiques...).
- La solution correspond aux labels **et** à la topologie du motif.
- Un nombre borné d'**insertions** et délétions peut-être autorisé



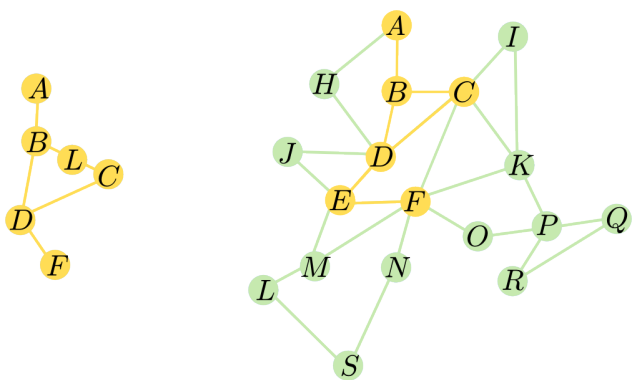
Recherche de motifs

- Vision "classique" de la recherche de motifs :
  - Le motif a une certaine topologie (chemin, arbre, graphes spécifiques...).
- La solution correspond aux labels **et** à la topologie du motif.
- Un nombre borné d'insertions et **délétions** peut-être autorisé



Recherche de motifs

- Vision "classique" de la recherche de motifs :
  - Le motif a une certaine topologie (chemin, arbre, graphes spécifiques...).
- La solution correspond aux labels **et** à la topologie du motif.
- Un nombre borné d'insertions et délétions peut-être autorisé
- Plusieurs d'algorithmes FPT, selon différents types de topologie [Shlomi et al. 2006, Dost et al. 2007] [Blin et al. 2009].



Graph Motif

- Constat : les données biologiques sont expérimentales donc **très bruitées** [Edwards et al. 2002]
  - Manque des informations (faux négatifs). Estimé à 50%.
  - Informations erronées (faux positifs). Estimé à 50%.
- La topologie du motif peut ne pas être connue *a priori*.
- Différents motifs peuvent avoir une même topologie.

Graph Motif

- Constat : les données biologiques sont expérimentales donc **très bruitées** [Edwards et al. 2002]
  - Manque des informations (faux négatifs). Estimé à 50%.
  - Informations erronées (faux positifs). Estimé à 50%.
- La topologie du motif peut ne pas être connue *a priori*.
- Différents motifs peuvent avoir une même topologie.
- La topologie du motif peut ne pas être pertinente.

Graph Motif – Définition

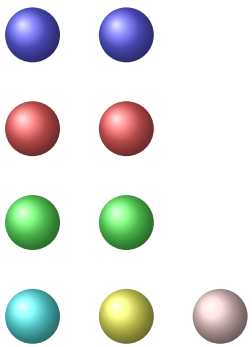
- Chaque nœud du réseau est coloré selon sa famille.
- Le motif est juste un (multi-)ensemble de couleurs à retrouver connecté dans le réseau coloré [Lacroix et al. 2006].

Graph Motif – Définition

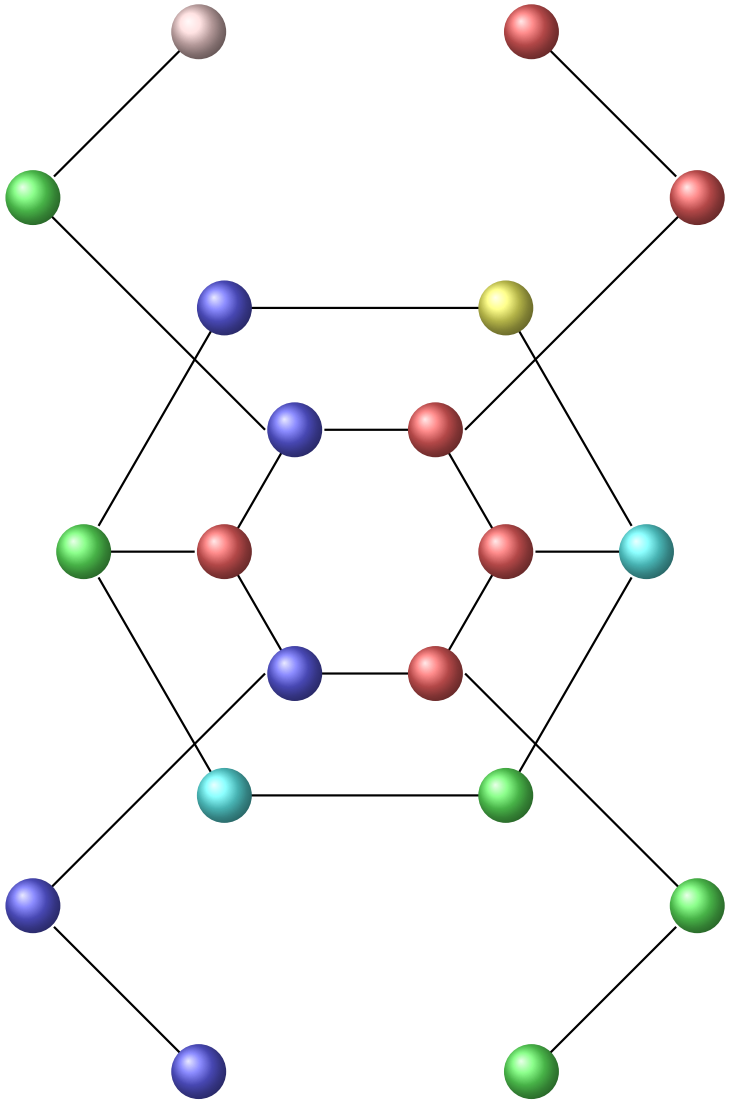
- Chaque nœud du réseau est coloré selon sa famille.
- Le motif est juste un (multi-)ensemble de couleurs à retrouver connecté dans le réseau coloré [Lacroix et al. 2006].
- Topologie n'est que la connexité de la solution.

Graph Motif – Un exemple jouet

M :

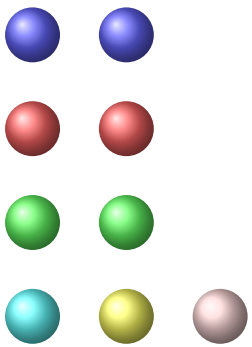


G :

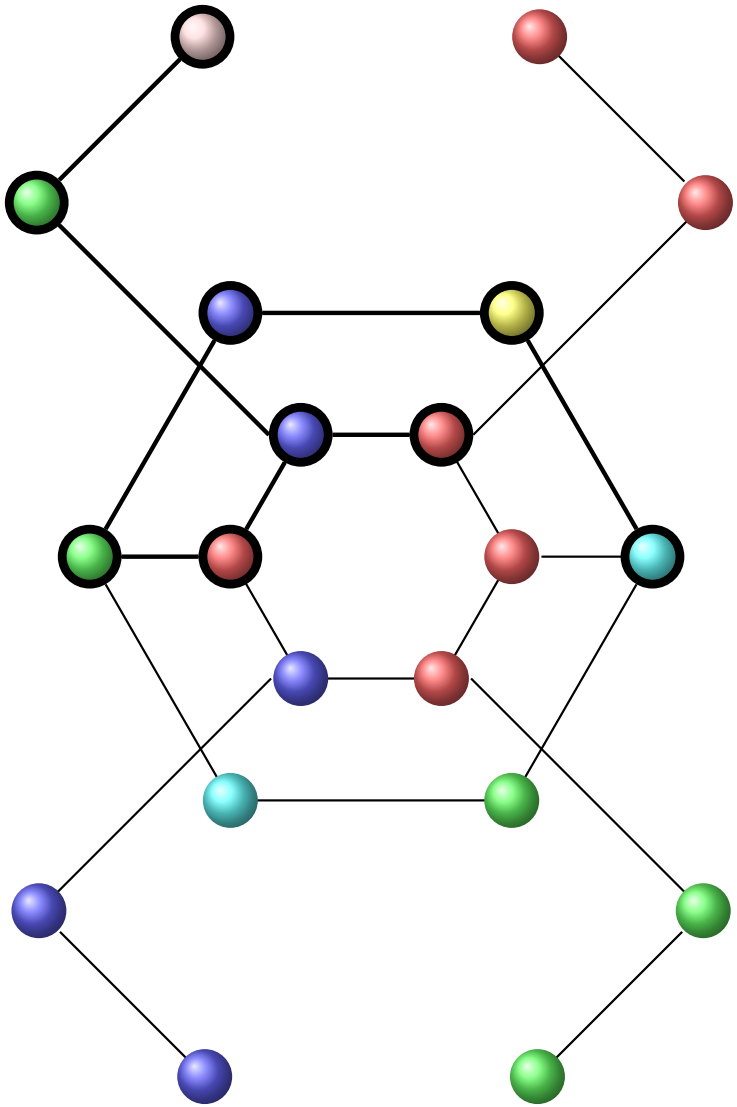


Graph Motif – Un exemple jouet

M :

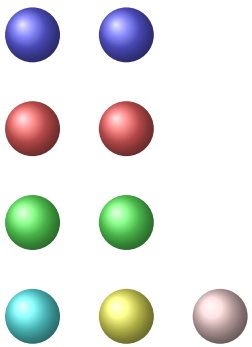


G :

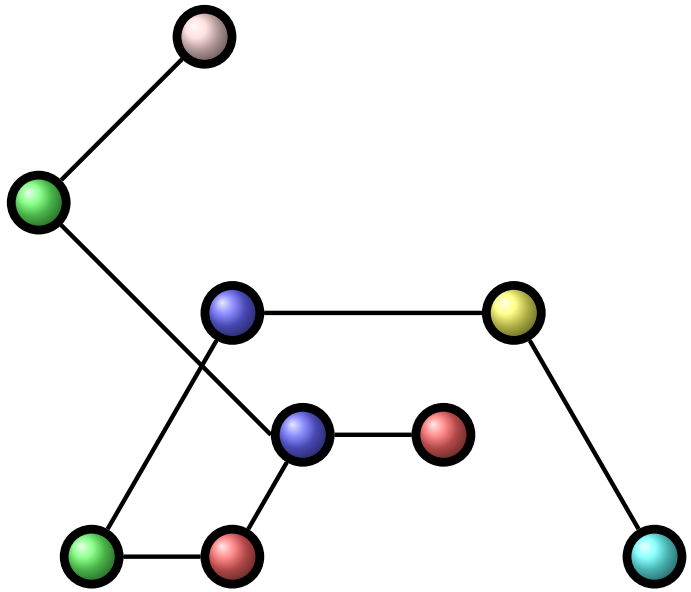


Graph Motif – Un exemple jouet

M :



G :



Graph Motif

- S'applique à différents types de réseaux biologiques.
  - Initialement sur les réseaux métaboliques [Lacroix et al. 2006].
  - Utilisable sur les réseaux PPI [Bruckner et al. 2009].
  - Applicable également à la spectrométrie de masse [Bocker et al. 2009].
- Selon [Betzler et al. 2008], peut être utilisé pour les réseaux sociaux.

Graph Motif – Difficulté

- Le problème est NP-complet, même si :
  - Le réseau est un arbre [Lacroix et al. 2006],
  - Cet **arbre est de degré maximum 3** et le motif est un **colorful** [Fellows et al. 2007],
  - L'**arbre enraciné est de profondeur 2** et le motif est un **colorful** [Ambalath et al. 2010],
  - Le motif n'est constitué que de **2 couleurs** et le réseau est un graphe biparti de degré maximum 4 [Fellows et al. 2007].

Graph Motif – Complexité paramétré

- Le problème est FPT par la taille k du motif :
  - $\mathcal{O}^*(87^k)$ [Fellows et al. 2007].
  - $\mathcal{O}^*(4.3^k)$ (motif multi-ensemble) [Betzler et al. 2008].
  - $\mathcal{O}^*(3^k)$ (motif colorful) [Betzler et al. 2008].
  - Pas de kernel polynomial, même si G est un *comb graph* [Ambalath et al. 2010].

Graph Motif – Complexité paramétré

- Le problème est FPT par la taille k du motif :
  - $\mathcal{O}^*(87^k)$ [Fellows et al. 2007].
  - $\mathcal{O}^*(4.3^k)$ (motif multi-ensemble) [Betzler et al. 2008].
  - $\mathcal{O}^*(3^k)$ (motif colorful) [Betzler et al. 2008].
  - Pas de kernel polynomial, même si G est un *comb graph* [Ambalath et al. 2010].
- Programmation dynamique et color-coding : **complexité en espace exponentielle** en k.

Graph Motif – Complexité paramétré

- Le problème est FPT par la taille k du motif :
  - $\mathcal{O}^*(87^k)$ [Fellows et al. 2007].
  - $\mathcal{O}^*(4.3^k)$ (motif multi-ensemble) [Betzler et al. 2008].
  - $\mathcal{O}^*(3^k)$ (motif colorful) [Betzler et al. 2008].
  - Pas de kernel polynomial, même si G est un *comb graph* [Ambalath et al. 2010].
- Programmation dynamique et color-coding : **complexité en espace exponentielle** en k.
- Mais W[1]-difficile si paramétré par le **nombre de couleurs** différentes, même si G est un arbre [Fellows et al. 2007].

Graph Motif et polynômes

- Amélioration de la complexité *via* un outil dû à Koutis et Williams [2008,2009].

Graph Motif et polynômes

- Amélioration de la complexité *via* un outil dû à Koutis et Williams [2008,2009].
- Résultat clef :
  - On peut déterminer par un algorithme randomisé, en **temps $\mathcal{O}^*(2^k)$ et en espace polynomial**, si un polynôme représenté par un circuit arithmétique contient un **monôme multilinéaire de degré k**.

Graph Motif et polynômes

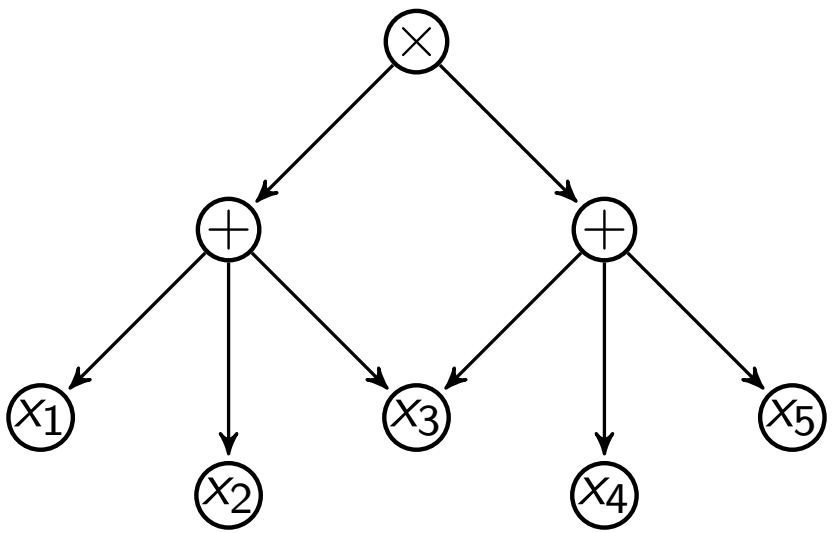
- Un monôme est **multilinéaire** si chaque variable du monôme n'apparait qu'une seule fois.
- Par définition, le degré d'un monôme multilinéaire est son nombre de variables.
- Exemple : $P(X) = (x_1^2 x_3 x_5 + x_1 x_2 x_4 x_6)$:
  - $x_1 x_2 x_4 x_6$ est un monôme multilinéaire de degré 4.
  - $x_1^2 x_3 x_5$ n'est pas un monôme multilinéaire.

Graph Motif et polynômes

- Un **circuit arithmétique** sur un ensemble de variables X est un DAG t.q. :
  - les noeuds internes sont les opérations $\times$ ou $+$,
  - les feuilles sont des éléments de X.

Graph Motif et polynômes

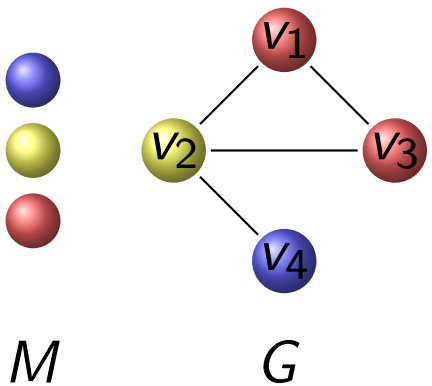
- Un **circuit arithmétique** sur un ensemble de variables X est un DAG t.q. :
  - les noeuds internes sont les opérations $\times$ ou $+$,
  - les feuilles sont des éléments de X.
- Exemple pour $P(X) = (x_1 + x_2 + x_3)(x_3 + x_4 + x_5)$.



Graph Motif et polynômes

- Pour résoudre Graph Motif avec **motif colorful** :
  - On introduit des **variables correspondant aux couleurs**.
  - On construit un circuit correspondant aux couleurs de **tous les sous-graphes (connexes) de taille** k de G.
  - Un monôme multilinéaire de degré k correspond alors à un sous-graphe avec k couleurs différentes (motif colorful), donc k sommets différents.

Exemple

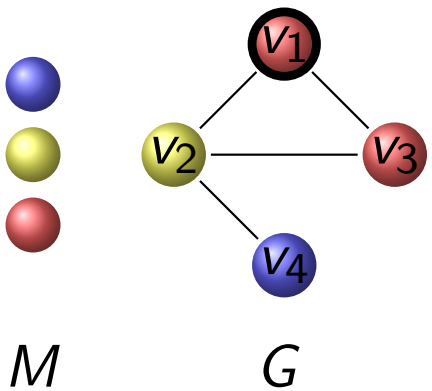


Construction récursive

Les monômes de degré $k = 3$ contenant le nœud v_1 sont la somme des produits des monômes :

1. de degrés $k' < k$ contenant le nœud v_1,
2. de degrés $k - k'$ contenant un voisin de v_1.

Exemple



Construction récursive

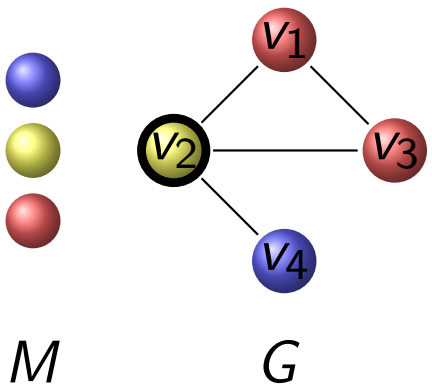
Les monômes de degré $k = 3$ contenant le nœud v_1 sont la somme des produits des monômes :

1. de degrés $k' < k$ contenant le nœud v_1,
2. de degrés $k - k'$ contenant un voisin de v_1.

Pour $k' = 1$:

x_R

Exemple



Construction récursive

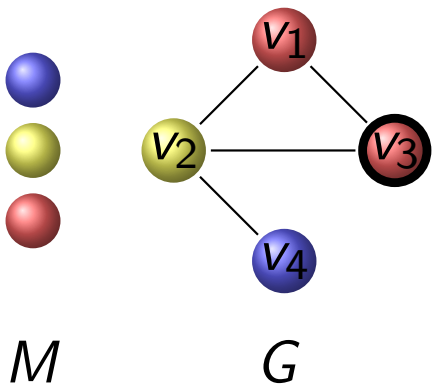
Les monômes de degré $k = 3$ contenant le nœud v_1 sont la somme des produits des monômes :

1. de degrés $k' < k$ contenant le nœud v_1,
2. de degrés $k - k'$ contenant un voisin de v_1.

Pour $k' = 1$:

$x_R \quad .(P_{2,v_2})$

Exemple



Construction récursive

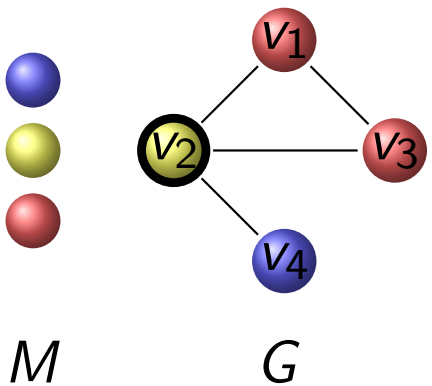
Les monômes de degré $k = 3$ contenant le nœud v_1 sont la somme des produits des monômes :

1. de degrés $k' < k$ contenant le nœud v_1,
2. de degrés $k - k'$ contenant un voisin de v_1.

Pour $k' = 1$:

$x_R \quad .(P_{2,v_2} + P_{2,v_3})$

Exemple



Construction récursive

Les monômes de degré $k = 3$ contenant le nœud v_1 sont la somme des produits des monômes :

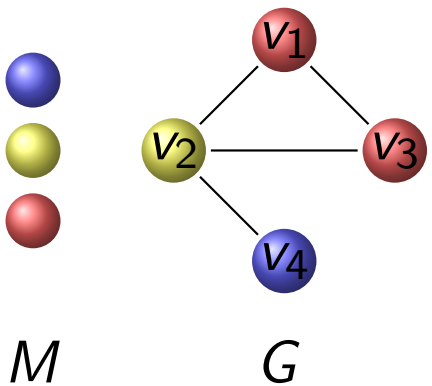
1. de degrés $k' < k$ contenant le nœud v_1,
2. de degrés $k - k'$ contenant un voisin de v_1.

Pour $k' = 1$:

$x_R \quad .(P_{2,v_2} + P_{2,v_3})$

$= x_R.(x_J.(P_{1,v_1} + P_{1,v_3} + P_{1,v_4}) + P_{2,v_3})$

Exemple



Construction récursive

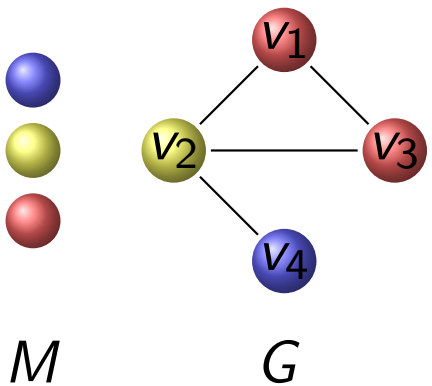
Les monômes de degré $k = 3$ contenant le nœud v_1 sont la somme des produits des monômes :

1. de degrés $k' < k$ contenant le nœud v_1,
2. de degrés $k - k'$ contenant un voisin de v_1.

Pour $k' = 1$:

$$x_R \quad .(P_{2,v_2}+P_{2,v_3})$$
$$= x_R.(x_J.(P_{1,v_1} + P_{1,v_3} + P_{1,v_4})+P_{2,v_3})$$
$$= x_R.(x_J.(x_R + x_R + x_B)+P_{2,v_3})$$

Exemple



Construction récursive

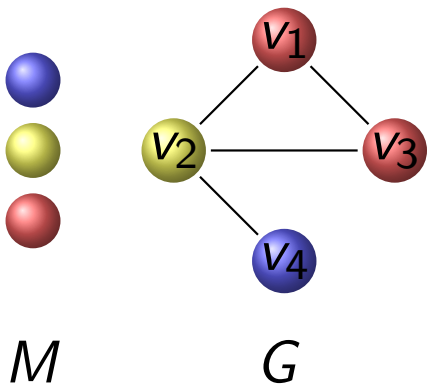
Les monômes de degré $k = 3$ contenant le nœud v_1 sont la somme des produits des monômes :

1. de degrés $k' < k$ contenant le nœud v_1,
2. de degrés $k - k'$ contenant un voisin de v_1.

Pour $k' = 1$:

$$\begin{aligned} & x_R \quad .(P_{2,v_2} + P_{2,v_3}) \\ &= x_R.(x_J.(P_{1,v_1} + P_{1,v_3} + P_{1,v_4}) + P_{2,v_3}) \\ &= x_R.(x_J.(x_R + x_R + x_B) + P_{2,v_3}) \\ &= x_R.(x_J.x_R + x_J.x_R + x_J.x_B + P_{2,v_3}) \end{aligned}$$

Exemple



Construction récursive

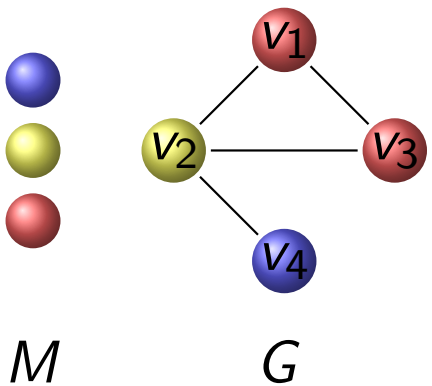
Les monômes de degré $k = 3$ contenant le nœud v_1 sont la somme des produits des monômes :

1. de degrés $k' < k$ contenant le nœud v_1,
2. de degrés $k - k'$ contenant un voisin de v_1.

Pour $k' = 1$:

$$
\begin{aligned}
& x_R \quad .(P_{2,v_2} + P_{2,v_3}) \\
&= x_R.(x_J.(P_{1,v_1} + P_{1,v_3} + P_{1,v_4}) + P_{2,v_3}) \\
&= x_R.(x_J.(x_R + x_R + x_B) + P_{2,v_3}) \\
&= x_R.(x_J.x_R + x_J.x_R + x_J.x_B + P_{2,v_3}) \\
&= x_R x_J x_R + x_R x_J x_R + x_R x_J x_B + \ldots
\end{aligned}
$$

Exemple



Construction récursive

Les monômes de degré $k = 3$ contenant le nœud v_1 sont la somme des produits des monômes :

1. de degrés $k' < k$ contenant le nœud v_1,
2. de degrés $k - k'$ contenant un voisin de v_1.

Pour $k' = 1$:

$$
\begin{aligned}
& x_R \quad .(P_{2,v_2} + P_{2,v_3}) \\
&= x_R.(x_J.(P_{1,v_1} + P_{1,v_3} + P_{1,v_4}) + P_{2,v_3}) \\
&= x_R.(x_J.(x_R + x_R + x_B) + P_{2,v_3}) \\
&= x_R.(x_J.x_R + x_J.x_R + x_J.x_B + P_{2,v_3}) \\
&= x_R x_J x_R + x_R x_J x_R + x_R x_J x_B + \ldots
\end{aligned}
$$

Il y a un monôme multilinéaire (donc une solution).

Graph Motif et polynômes

- La construction est polynomiale.
- Il existe **un monôme multilinéaire de degré k** si et seulement s'il existe **un sous-graphe connexe colorful de taille k**.

Graph Motif et polynômes

- La construction est polynomiale.
- Il existe **un monôme multilinéaire de degré k** si et seulement s'il existe **un sous-graphe connexe colorful de taille k**.

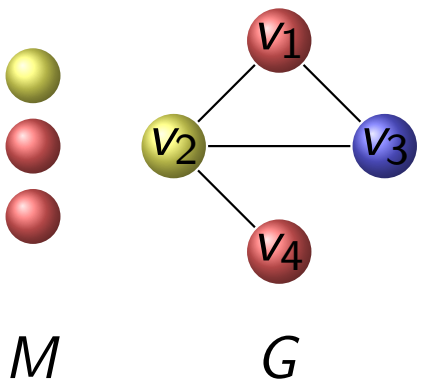
- On peut résoudre Graph Motif lorsque le motif est colorful en $\mathcal{O}^*(2^k)$ (algorithme randomisé) et en espace polynomial [Guillemot et S. 2011].

Graph Motif et polynômes – Multi-ensemble

- Si le **motif est un multi-ensemble**, un monôme non multilinéaire sur les couleurs est une solution.
- La construction précédente n'est plus suffisante.

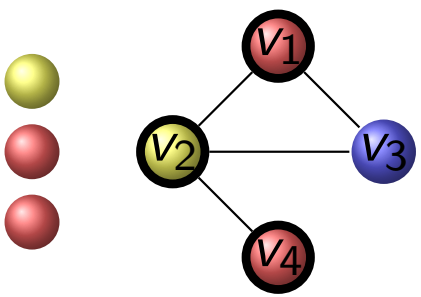
Graph Motif et polynômes – Multi-ensemble

- Si le **motif est un multi-ensemble**, un monôme non multilinéaire sur les couleurs est une solution.
- La construction précédente n'est plus suffisante.



Graph Motif et polynômes – Multi-ensemble

- Si le **motif est un multi-ensemble**, un monôme non multilinéaire sur les couleurs est une solution.
- La construction précédente n'est plus suffisante.

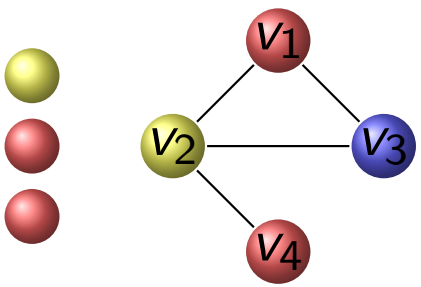


M G

$$\cdots = x_R x_J x_R + x_R x_J x_B + x_R x_J x_R + \ldots$$

Graph Motif et polynômes – Multi-ensemble

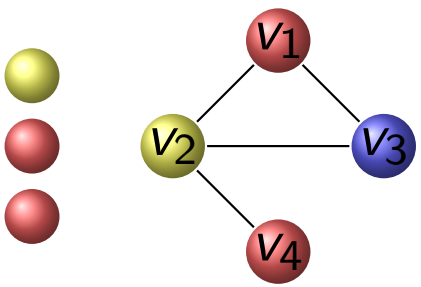
- Si le **motif est un multi-ensemble**, un monôme non multilinéaire sur les couleurs est une solution.
- La construction précédente n'est plus suffisante.
- Pour chaque couleur c qui apparait m fois dans le motif, on introduit des variables $x_{c,1}, x_{c,2}, \ldots, x_{c,m}$.



M G

Graph Motif et polynômes – Multi-ensemble

- Si le **motif est un multi-ensemble**, un monôme non multilinéaire sur les couleurs est une solution.
- La construction précédente n'est plus suffisante.
- Pour chaque couleur c qui apparait m fois dans le motif, on introduit des variables $x_{c,1}, x_{c,2}, \ldots, x_{c,m}$.

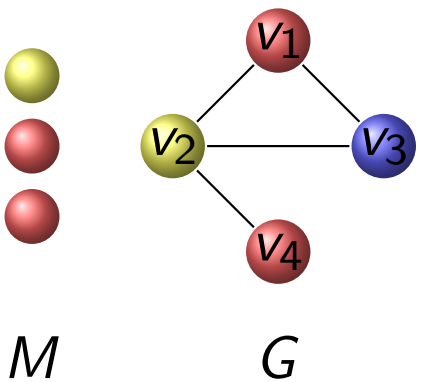


M G

$$\cdots = (x_{R,1} + x_{R,2})x_{J,1}(x_{R,1} + x_{R,2}) + \ldots$$

Graph Motif et polynômes – Multi-ensemble

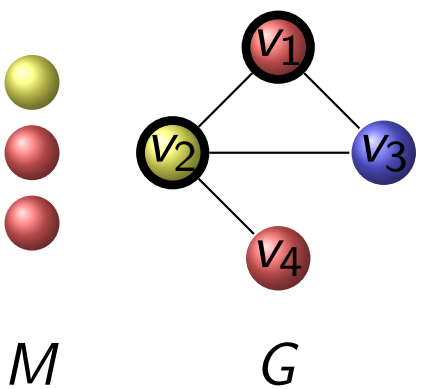
- Si le **motif est un multi-ensemble**, un monôme non multilinéaire sur les couleurs est une solution.
- La construction précédente n'est plus suffisante.
- Pour chaque couleur c qui apparait m fois dans le motif, on introduit des variables $x_{c,1}, x_{c,2}, \ldots, x_{c,m}$ (reste insuffisant).



$$\cdots = (x_{R,1} + x_{R,2})x_{J,1}(x_{R,1} + x_{R,2}) + \ldots$$
$$= x_{R,1}x_{J,1}x_{R,1} + x_{R,1}x_{J,1}x_{R,2} + x_{R,2}x_{J,1}x_{R,1}$$
$$+ x_{R,2}x_{J,1}x_{R,2} + \ldots$$

Graph Motif et polynômes – Multi-ensemble

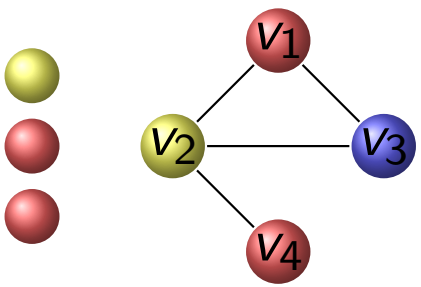
- Si le **motif est un multi-ensemble**, un monôme non multilinéaire sur les couleurs est une solution.
- La construction précédente n'est plus suffisante.
- Pour chaque couleur c qui apparait m fois dans le motif, on introduit des variables $x_{c,1}, x_{c,2}, \ldots, x_{c,m}$ (reste insuffisant).



$$\cdots = (x_{R,1} + x_{R,2})x_{J,1}(x_{R,1} + x_{R,2}) + \ldots$$
$$= x_{R,1}x_{J,1}x_{R,1} + x_{R,1}x_{J,1}x_{R,2} + x_{R,2}x_{J,1}x_{R,1}$$
$$+ x_{R,2}x_{J,1}x_{R,2} + \ldots$$

Graph Motif et polynômes – Multi-ensemble

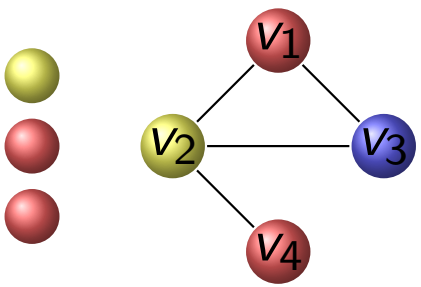
- Si le **motif est un multi-ensemble**, un monôme non multilinéaire sur les couleurs est une solution.
- La construction précédente n'est plus suffisante.
- Pour chaque couleur c qui apparait m fois dans le motif, on introduit des variables $x_{c,1}, x_{c,2}, \ldots, x_{c,m}$ (reste insuffisant).
- Ajout de variables y_v pour les nœuds v du graphe.



M G

Graph Motif et polynômes – Multi-ensemble

- Si le **motif est un multi-ensemble**, un monôme non multilinéaire sur les couleurs est une solution.
- La construction précédente n'est plus suffisante.
- Pour chaque couleur c qui apparait m fois dans le motif, on introduit des variables $x_{c,1}, x_{c,2}, \ldots, x_{c,m}$ (reste insuffisant).
- Ajout de variables y_v pour les nœuds v du graphe.

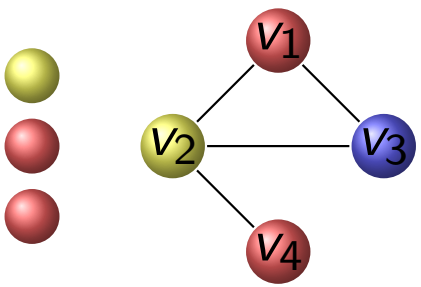


M G

$$\cdots = y_1(x_{R,1} + x_{R,2}) \cdot y_2 x_{J,1} \cdot y_1(x_{R,1} + x_{R,2}) + \ldots$$

Graph Motif et polynômes – Multi-ensemble

- Si le **motif est un multi-ensemble**, un monôme non multilinéaire sur les couleurs est une solution.
- La construction précédente n'est plus suffisante.
- Pour chaque couleur c qui apparait m fois dans le motif, on introduit des variables $x_{c,1}, x_{c,2}, \ldots, x_{c,m}$ (reste insuffisant).
- Ajout de variables y_v pour les nœuds v du graphe.

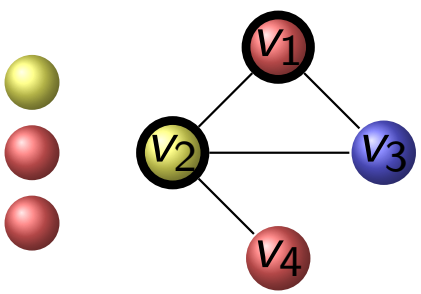


M G

$$\cdots = y_1(x_{R,1} + x_{R,2}) \cdot y_2 x_{J,1} \cdot y_1(x_{R,1} + x_{R,2}) + \ldots$$
$$= y_1 x_{R,1} y_2 x_{J,1} y_1 x_{R,1} + y_1 x_{R,1} y_2 x_{J,1} y_1 x_{R,2} + \ldots$$

Graph Motif et polynômes – Multi-ensemble

- Si le **motif est un multi-ensemble**, un monôme non multilinéaire sur les couleurs est une solution.
- La construction précédente n'est plus suffisante.
- Pour chaque couleur c qui apparait m fois dans le motif, on introduit des variables $x_{c,1}, x_{c,2}, \ldots, x_{c,m}$ (reste insuffisant).
- Ajout de variables y_v pour les nœuds v du graphe.



M G

$$\cdots = y_1(x_{R,1} + x_{R,2}) \cdot y_2 x_{J,1} \cdot y_1(x_{R,1} + x_{R,2}) + \ldots$$
$$= y_1 x_{R,1} y_2 x_{J,1} y_1 x_{R,1} + y_1 x_{R,1} y_2 x_{J,1} y_1 x_{R,2} + \ldots$$

Graph Motif et polynômes – Multi-ensemble

- La construction est toujours polynomiale.
- Il existe un **monôme multilinéaire de degré** $2k$ si et seulement si il existe une **solution de taille** k.

Graph Motif et polynômes – Multi-ensemble

- La construction est toujours polynomiale.
- Il existe un **monôme multilinéaire de degré** $2k$ si et seulement si il existe une **solution de taille** k.

- On peut résoudre Graph Motif lorsque le motif est un multi-ensemble en $\mathcal{O}^*(4^k)$ (algorithme randomisé) et en espace polynomial [Guillemot et S. 2011].

- Implémentation comparable à l'état de l'art en pratique.

Graph Motif et polynômes – La suite

- Algorithme randomisé en $\mathcal{O}^*(2^k)$ et espace polynomial, même dans le cas multi-ensemble [Kowalik, pre-print 2012].
- Approche par Narrow sieves [Bjorklund et al., 2012].

Graph Motif et polynômes – La suite

- Algorithme randomisé en $\mathcal{O}^*(2^k)$ et espace polynomial, même dans le cas multi-ensemble [Kowalik, pre-print 2012].
- Approche par Narrow sieves [Bjorklund et al., 2012].
- Un algorithme en $\mathcal{O}((2-\epsilon)^k)$ (même pour le cas ensemble simple) impliquerait un algorithme en $\mathcal{O}((2-\epsilon')^n)$ pour Set Cover (ce qui est peu crédible (SETH) [Cygan et al., 2012]) [Kowalik, pre-print 2012].

Graph Motif et polynômes

- La méthode précédente répond au problème de décision.
- On peut **construire** une solution au prix d'un facteur $|V|$.

Graph Motif et polynômes

- La méthode précédente répond au problème de décision.
- On peut **construire** une solution au prix d'un facteur $|V|$.
- Pour chaque sommet v, tester si une solution de taille k existe dans le graphe $G = G \setminus \{v\}$:
  - Si oui : continuer.
  - Si non : remettre le sommet v.
- Le graphe final correspond à une solution.

Graph Motif et polynômes – Comptage

- **Compter** le nombre d'occurrences d'un motif **colorful** : $\mathcal{O}^*(2^k)$ (construction similaire).

Graph Motif et polynômes – Comptage

- **Compter** le nombre d'occurrences d'un motif **colorful** : $\mathcal{O}^*(2^k)$ (construction similaire).
- Mais, si le motif est un **multi-ensemble**, le problème est #W[1]-difficile [Guillemot et S. 2011].

Plan

Introduction

Complexité paramétrée

Recherche exacte de motifs

Recherche approchée de motifs

Perspectives

Variantes

- Données expérimentales, **imprécisions**.
- Rechercher une occurrence exacte d'un motif peut échouer.

Variantes

- Données expérimentales, **imprécisions**.
- Rechercher une occurrence exacte d'un motif peut échouer.

- Autoriser des sommets supplémentaires dans la solution pour atteindre la connexité (**insertions**).

Variantes

- Données expérimentales, **imprécisions**.
- Rechercher une occurrence exacte d'un motif peut échouer.

- Autoriser des sommets supplémentaires dans la solution pour atteindre la connexité (**insertions**).
- Autoriser l'absence de couleurs du motif dans la solution (**délétions**).

Variantes

- Données expérimentales, **imprécisions**.
- Rechercher une occurrence exacte d'un motif peut échouer.

- Autoriser des sommets supplémentaires dans la solution pour atteindre la connexité (**insertions**).
- Autoriser l'absence de couleurs du motif dans la solution (**délétions**).
- Autoriser **plusieurs composantes connexes** dans la solution.

Variantes

- Données expérimentales, **imprécisions**.
- Rechercher une occurrence exacte d'un motif peut échouer.

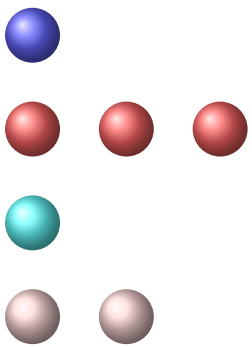
- Autoriser des sommets supplémentaires dans la solution pour atteindre la connexité (**insertions**).
- Autoriser l'absence de couleurs du motif dans la solution (**délétions**).
- Autoriser **plusieurs composantes connexes** dans la solution.
- Autoriser une **liste de couleurs** pour chaque sommet du réseau.

Une variante de Graph Motif : Minimum Substitutions

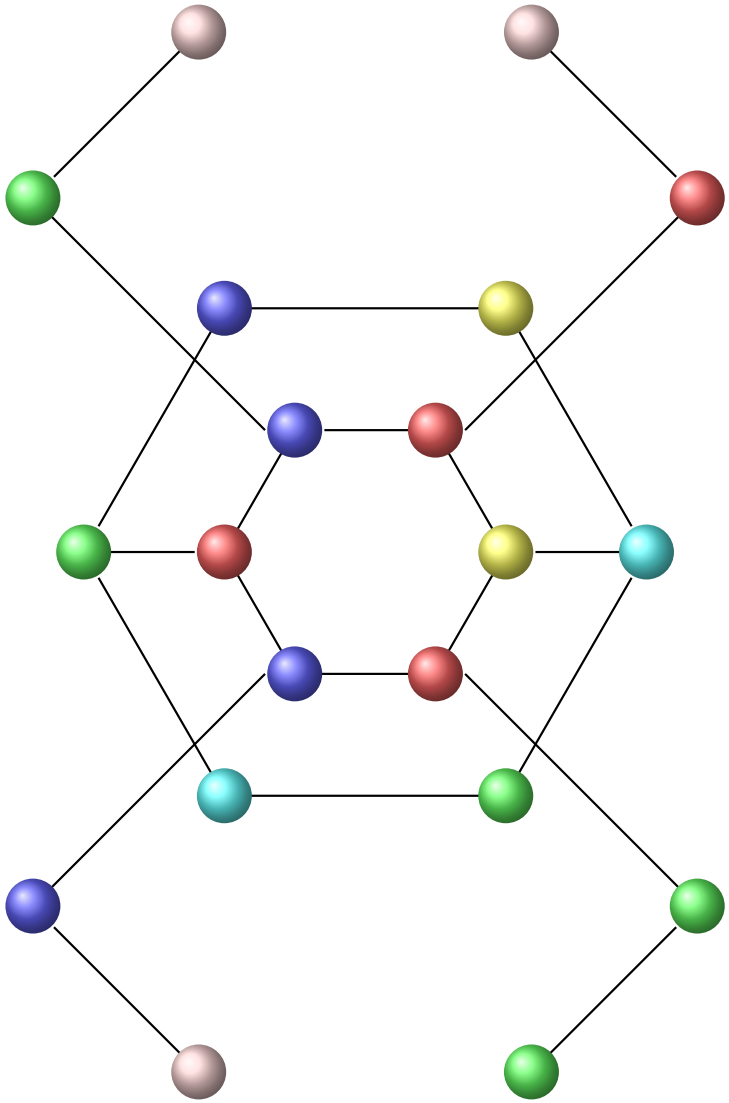
- Une autre variante : Minimum Substitutions [Dondi et al. 2011].
- Trouver une occurrence qui reprend "le plus de couleurs possible" du motif, mais de même taille que le motif.
- On *substitue* des couleurs du motif par de nouvelles couleurs.

Minimum Substitutions : Exemple

M :

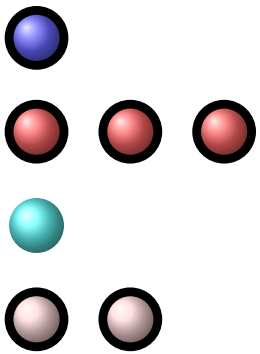


G :

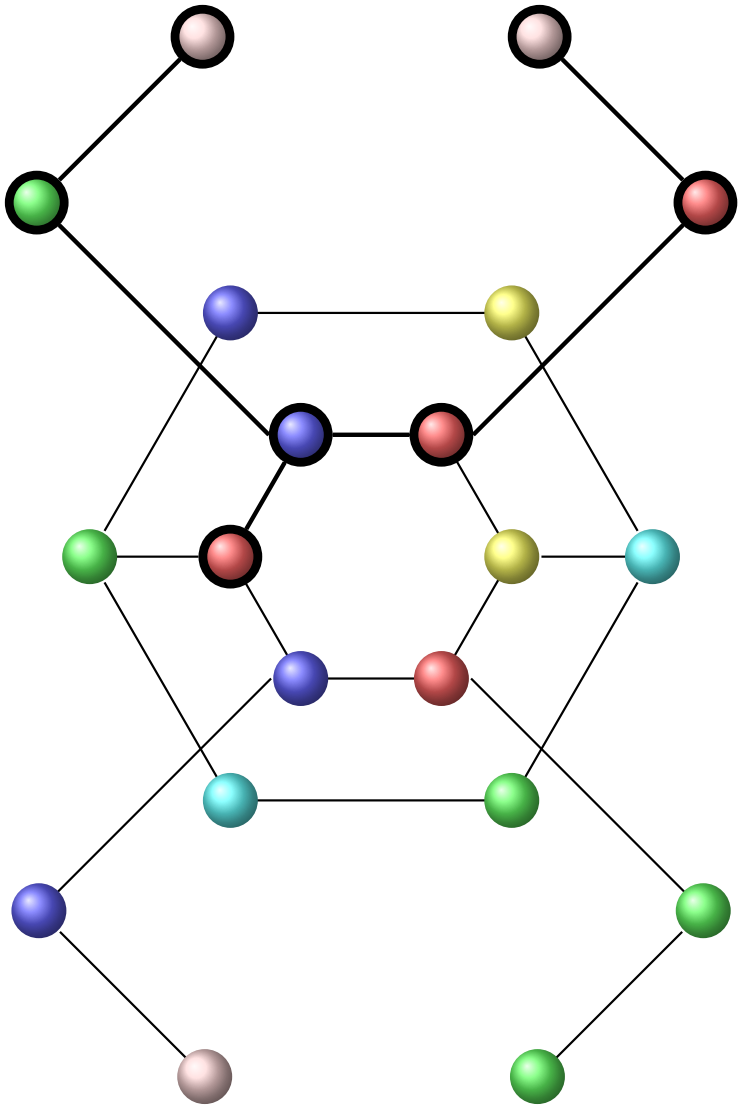


Minimum Substitutions : Exemple

M :

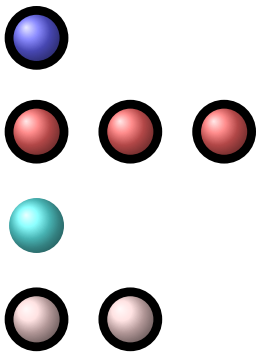


G :

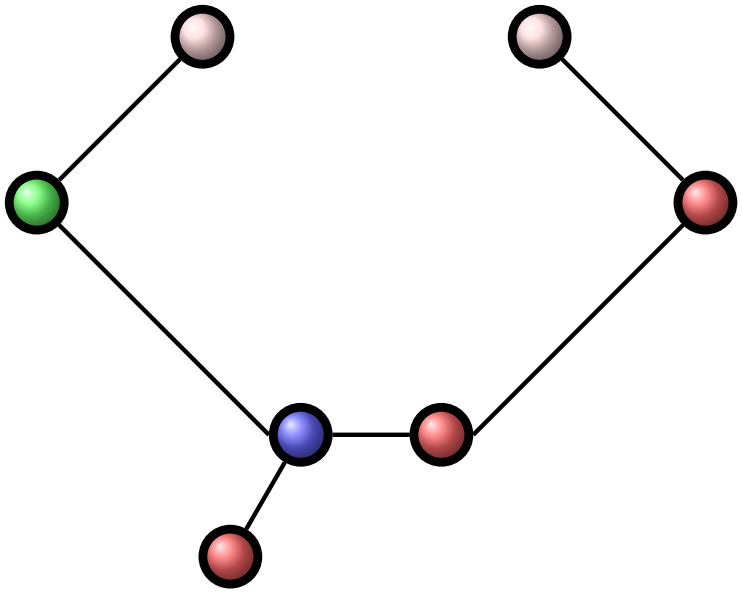


Minimum Substitutions : Exemple

M :



G :



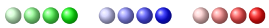
Une variante de Graph Motif : Minimum Substitutions

- Il n'existe pas de ratio d'approximation inférieur à $c \log |V|$, sauf si P = NP, même lorsque le motif est colorful et G est un arbre de profondeur 2.

Construction

$$X = \{x_1, x_2, x_3\}, \mathcal{S} = \{\{x_1, x_2\},\{x_2, x_3\},\{x_2\}\}$$

Construction



$$X = \{x_1, x_2, x_3\}, \mathcal{S} = \{\{x_1, x_2\}, \{x_2, x_3\}, \{x_2\}\}$$

Construction

$$X = \{x_1, x_2, x_3\}, \mathcal{S} = \{\{x_1, x_2\}, \{x_2, x_3\}, \{x_2\}\}$$

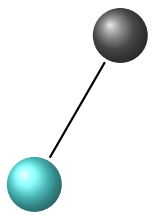
Construction

$$X = \{x_1, x_2, x_3\}, \mathcal{S} = \{\{x_1, x_2\}, \{x_2, x_3\}, \{x_2\}\}$$

M G

Construction

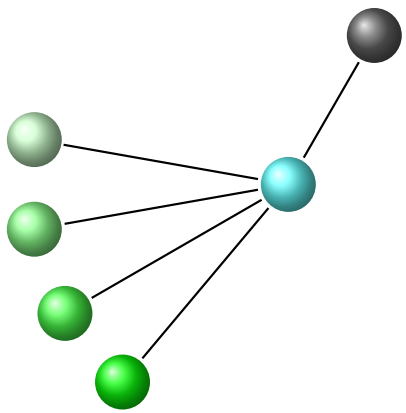
$$X = \{x_1, x_2, x_3\}, \mathcal{S} = \{\{x_1, x_2\}, \{x_2, x_3\}, \{x_2\}\}$$



M G

Construction

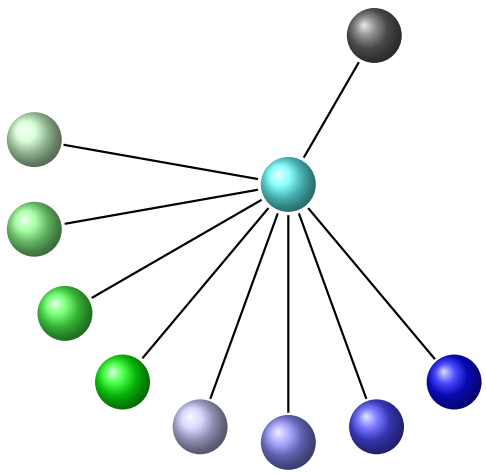
$$X = \{x_1, x_2, x_3\}, \mathcal{S} = \{\{x_1, x_2\}, \{x_2, x_3\}, \{x_2\}\}$$



M G

Construction

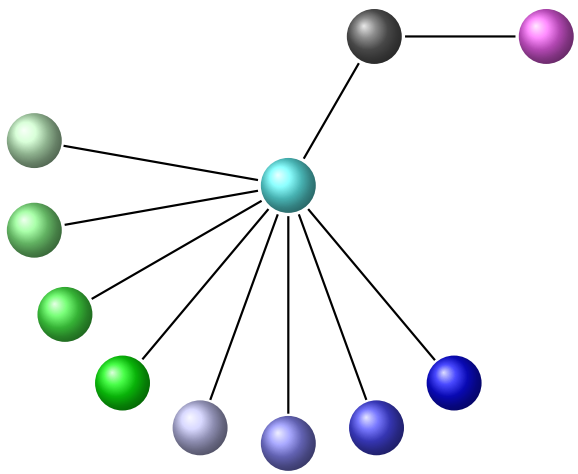
$$X = \{x_1, x_2, x_3\}, \mathcal{S} = \{\{x_1, x_2\}, \{x_2, x_3\}, \{x_2\}\}$$



M G

Construction

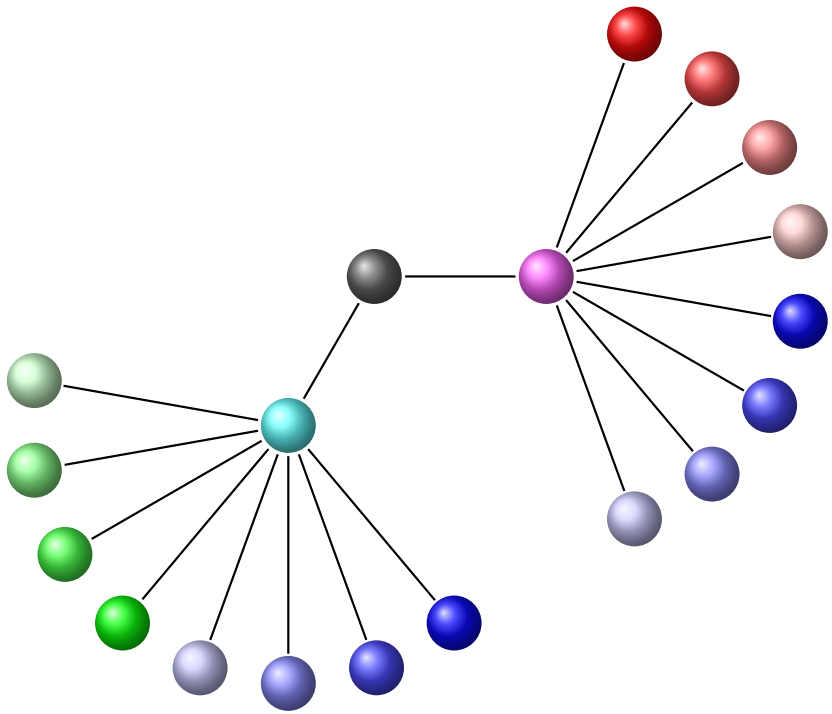
$$X = \{x_1, x_2, x_3\}, \mathcal{S} = \{\{x_1, x_2\}, \{x_2, x_3\}, \{x_2\}\}$$



M G

Construction

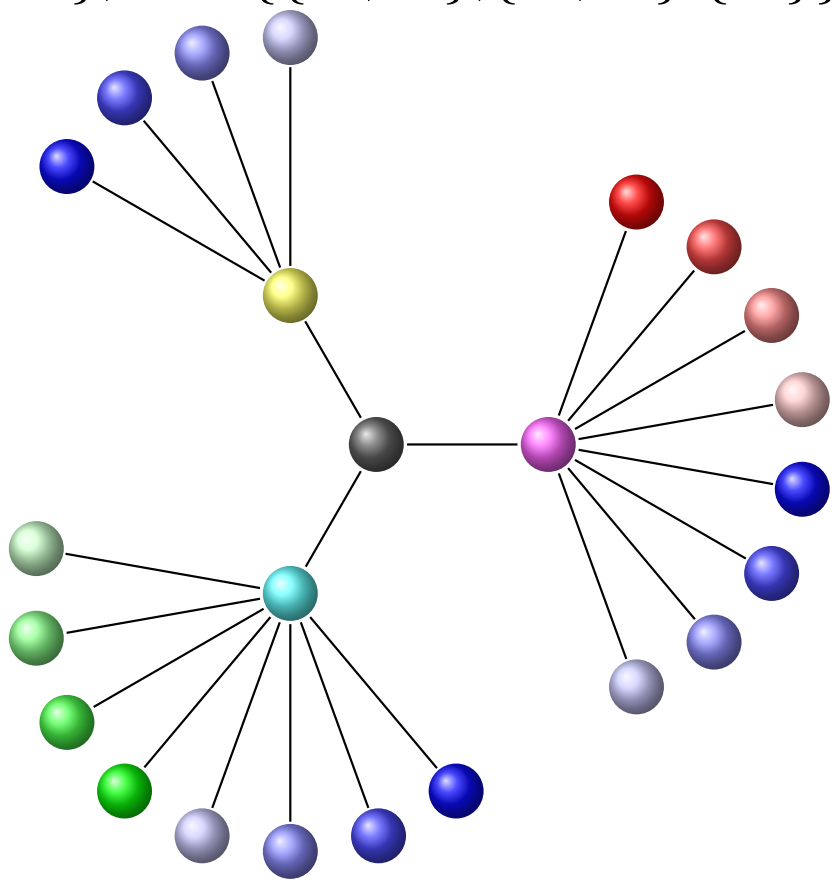
$$X = \{x_1, x_2, x_3\}, \mathcal{S} = \{\{x_1, x_2\}, \{x_2, x_3\}, \{x_2\}\}$$



M G

Construction

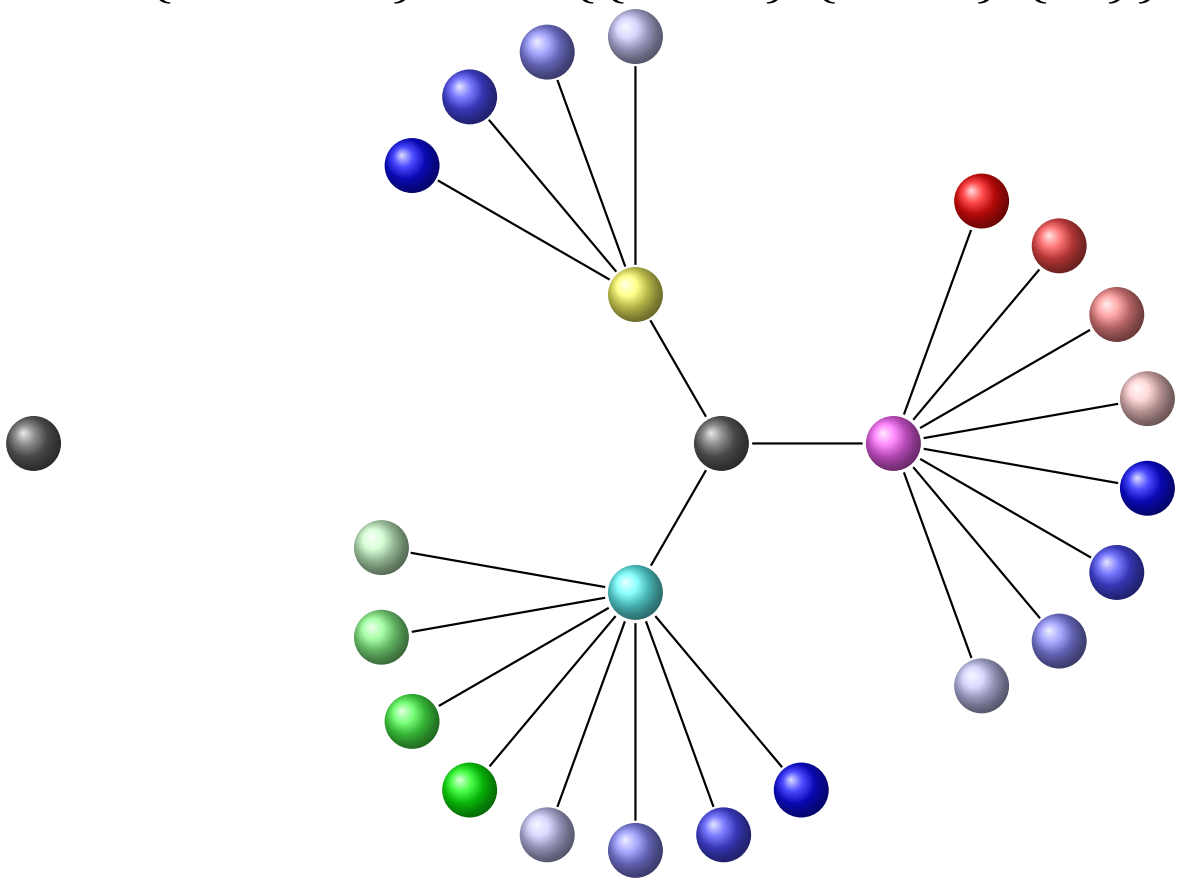
$$X = \{x_1, x_2, x_3\}, \mathcal{S} = \{\{x_1, x_2\}, \{x_2, x_3\}, \{x_2\}\}$$



M G

Construction

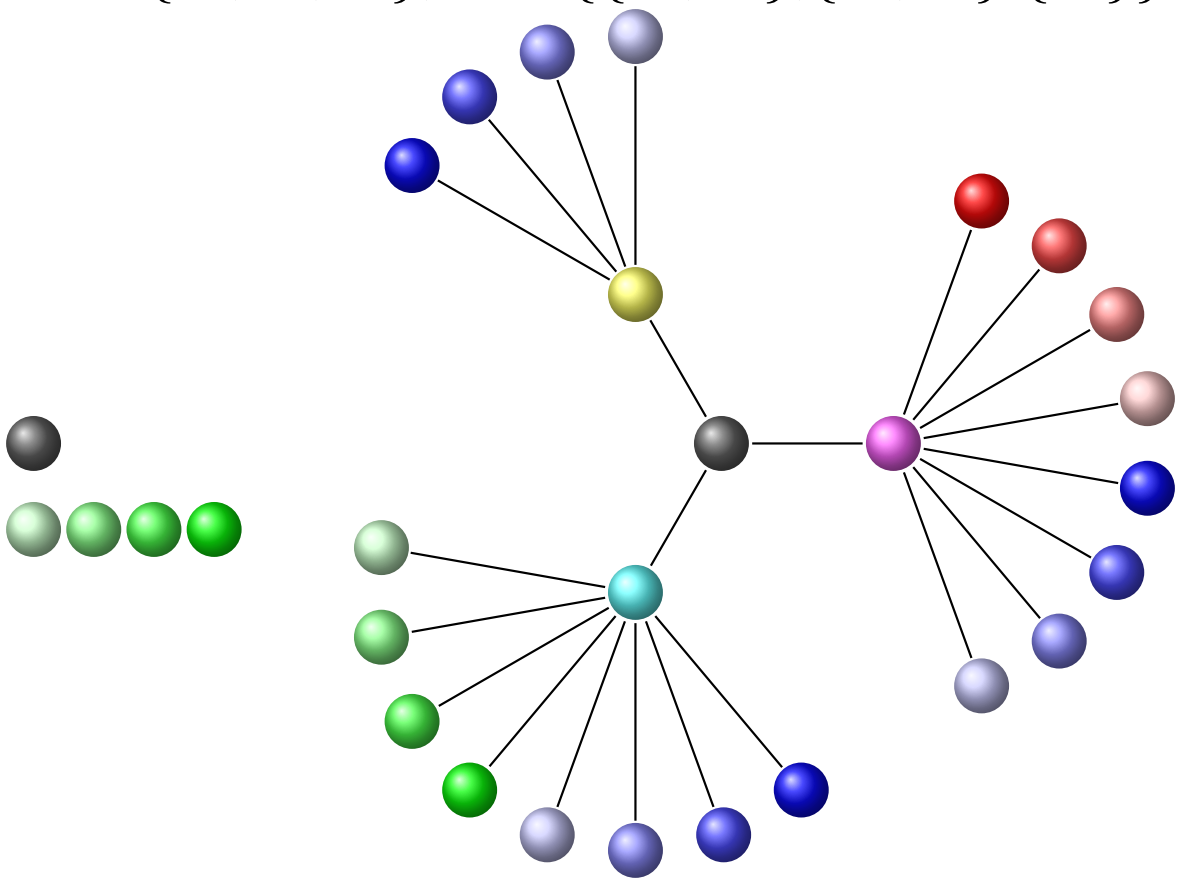
$$X = \{x_1, x_2, x_3\}, \mathcal{S} = \{\{x_1, x_2\}, \{x_2, x_3\}, \{x_2\}\}$$



M G

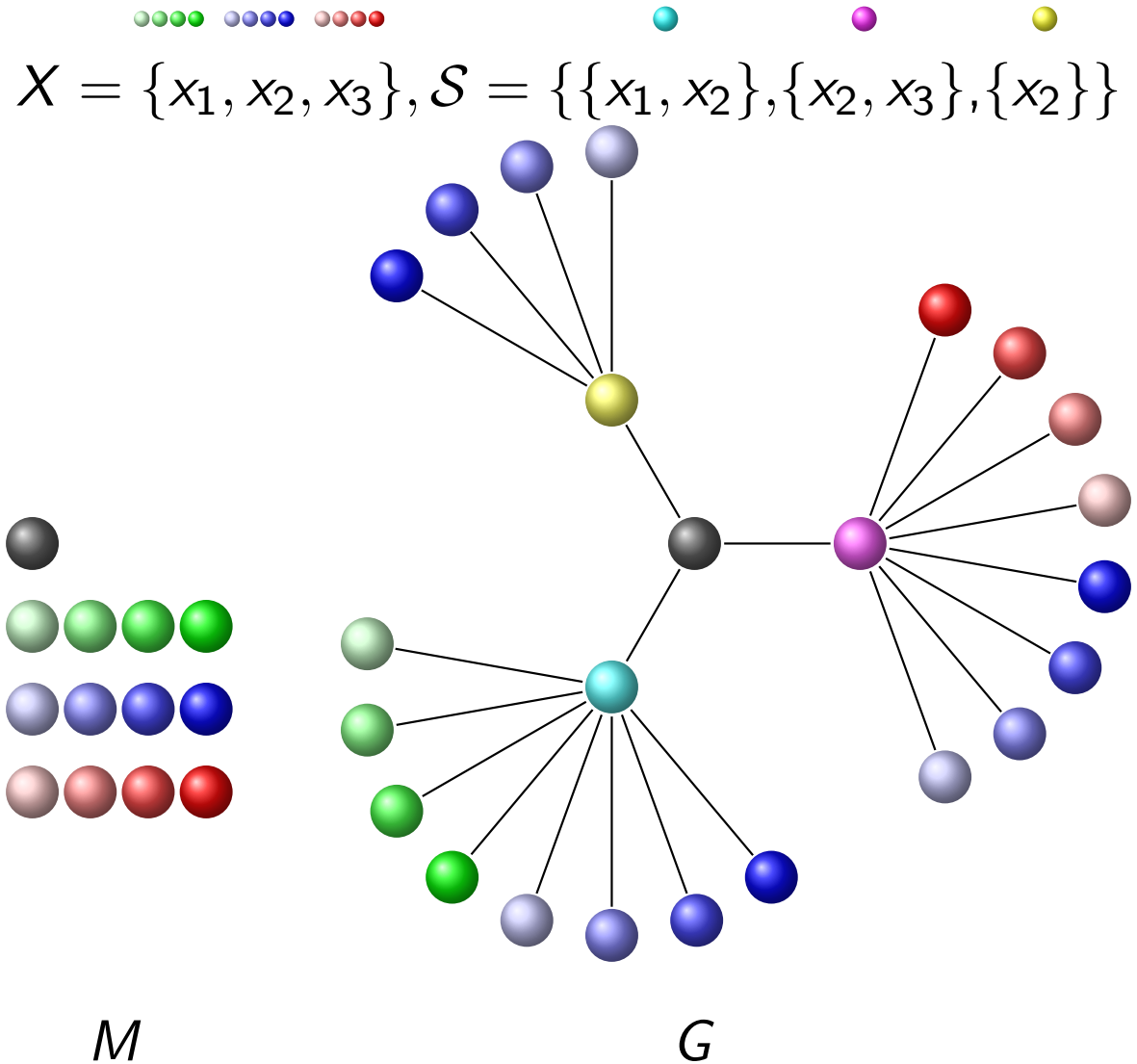
Construction

$$X = \{x_1, x_2, x_3\}, \mathcal{S} = \{\{x_1, x_2\}, \{x_2, x_3\}, \{x_2\}\}$$



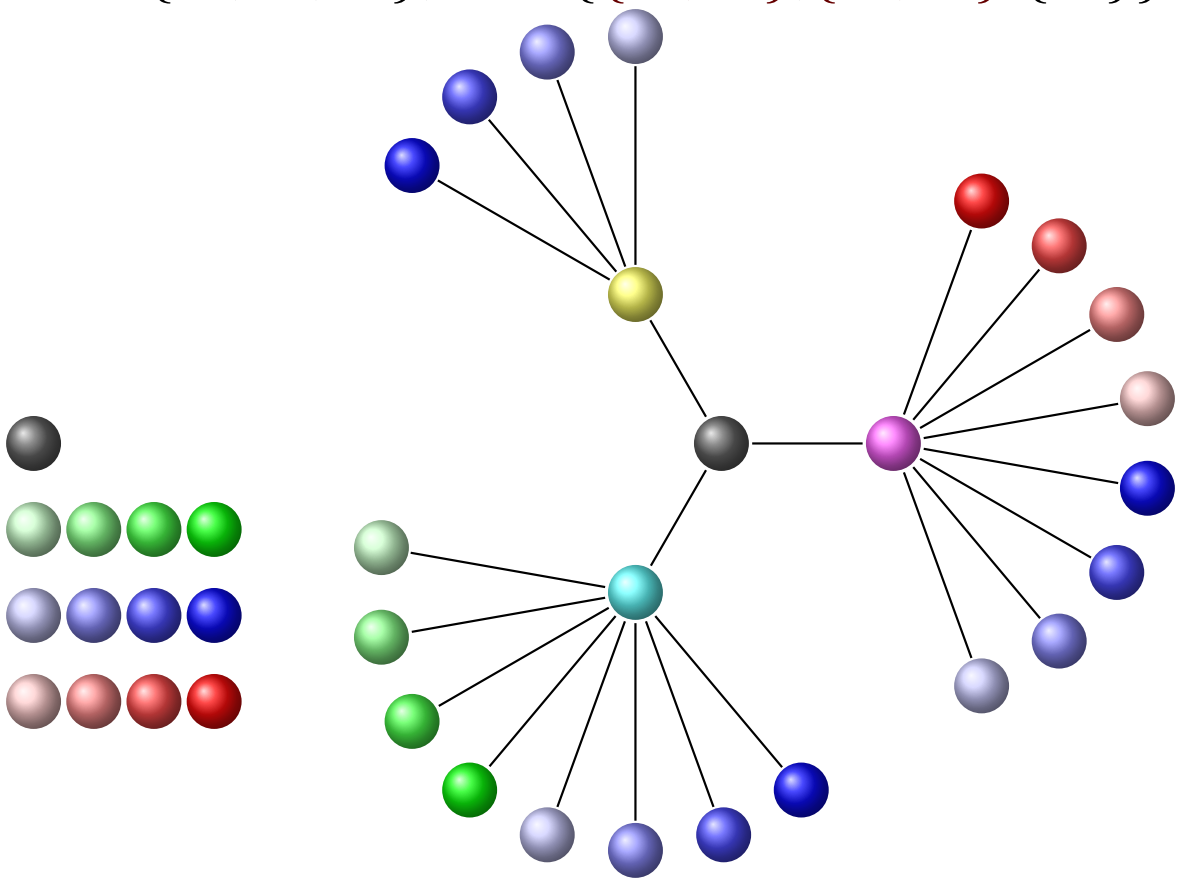
M G

Construction



Construction

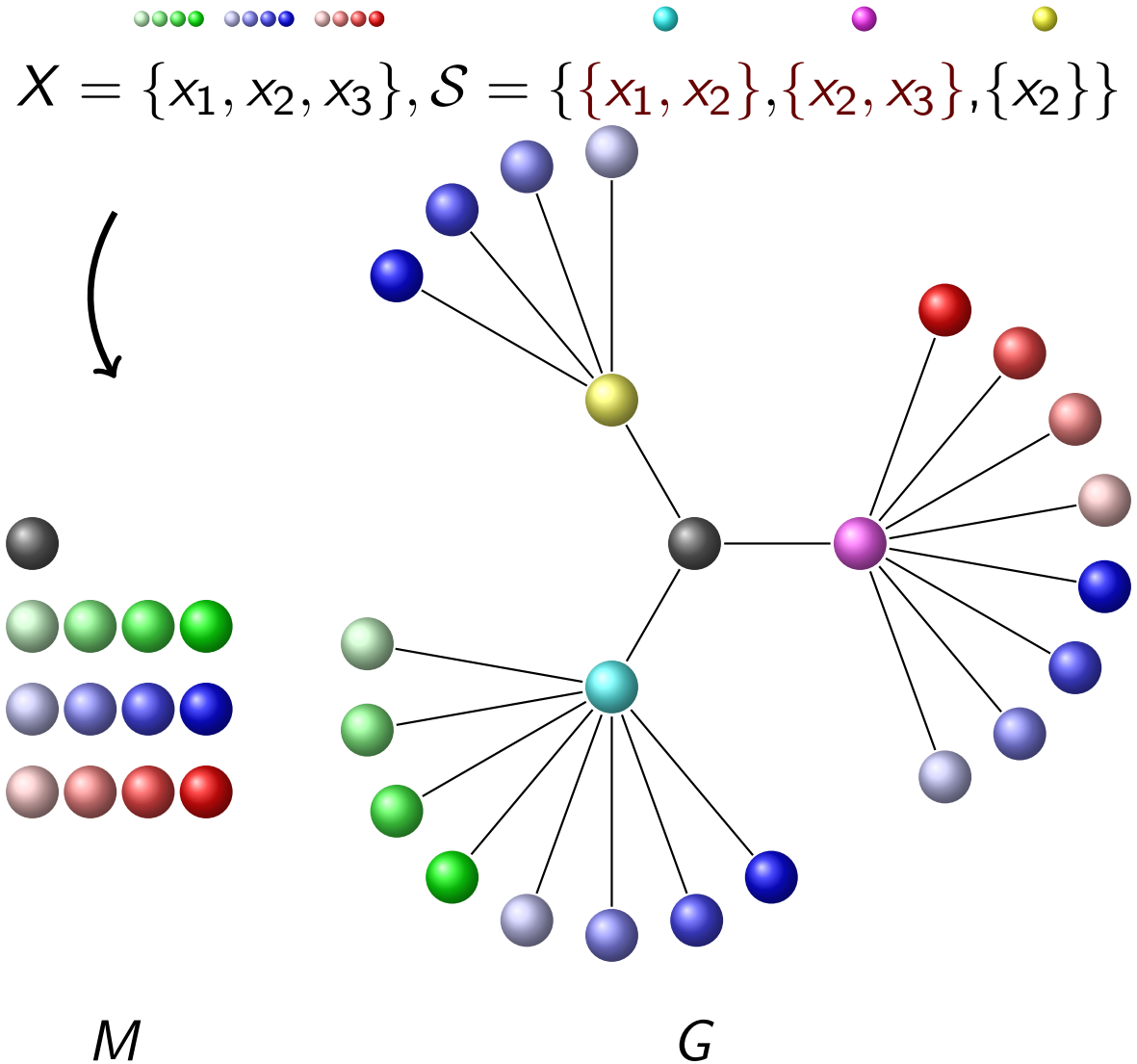
$X = \{x_1, x_2, x_3\}, \mathcal{S} = \{\{x_1, x_2\}, \{x_2, x_3\}, \{x_2\}\}$



M G

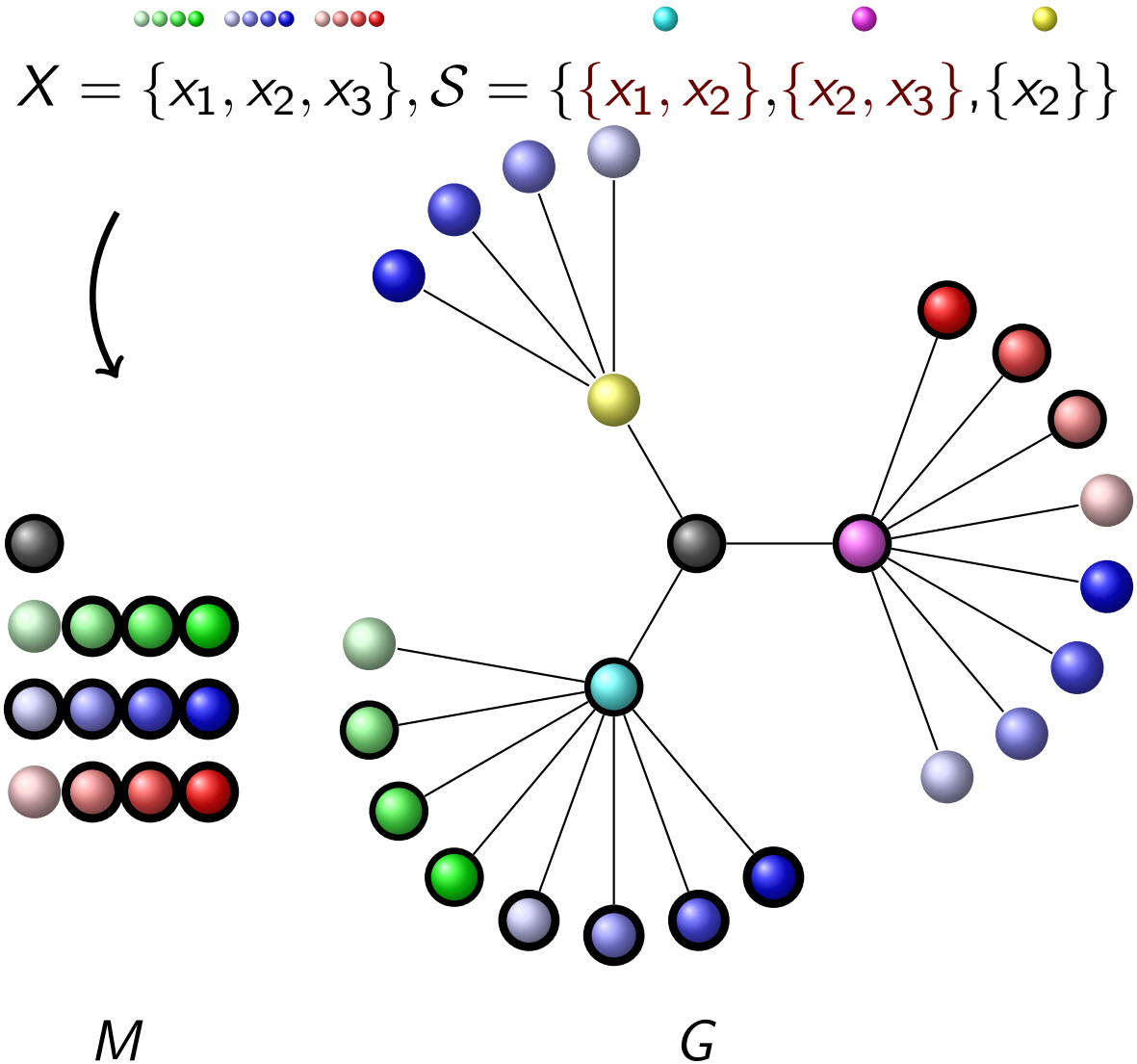
$|\mathcal{S}'|$ = nombre de substitutions

Construction



$|\mathcal{S}'|$ = nombre de substitutions

Construction



$|\mathcal{S}'|$ = nombre de substitutions

Résultat

- Il n'existe pas de ratio d'approximation meilleur que $c \log |X|$ pour Minimum Set Cover (sauf si P = NP) [Raz et Safra 1997].

Résultat

- Il n'existe pas de ratio d'approximation meilleur que $c \log |X|$ pour Minimum Set Cover (sauf si P = NP) [Raz et Safra 1997].
- Il n'existe pas de ratio d'approximation meilleur que $c \log |V|$ pour Minimum Substitutions (sauf si P = NP) [Rizzi et S. 2012].

Plan

Introduction

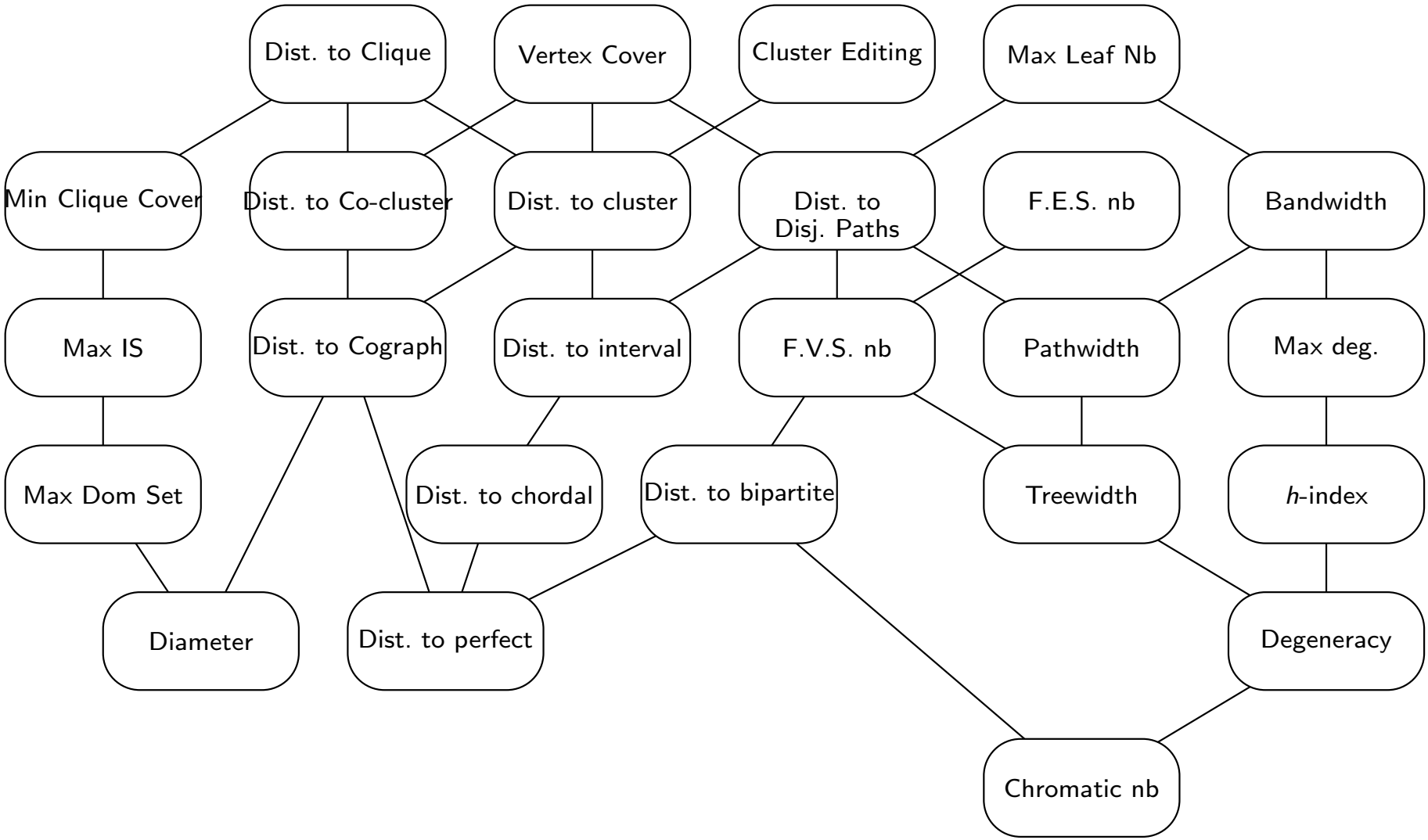
Complexité paramétrée

Recherche exacte de motifs

Recherche approchée de motifs

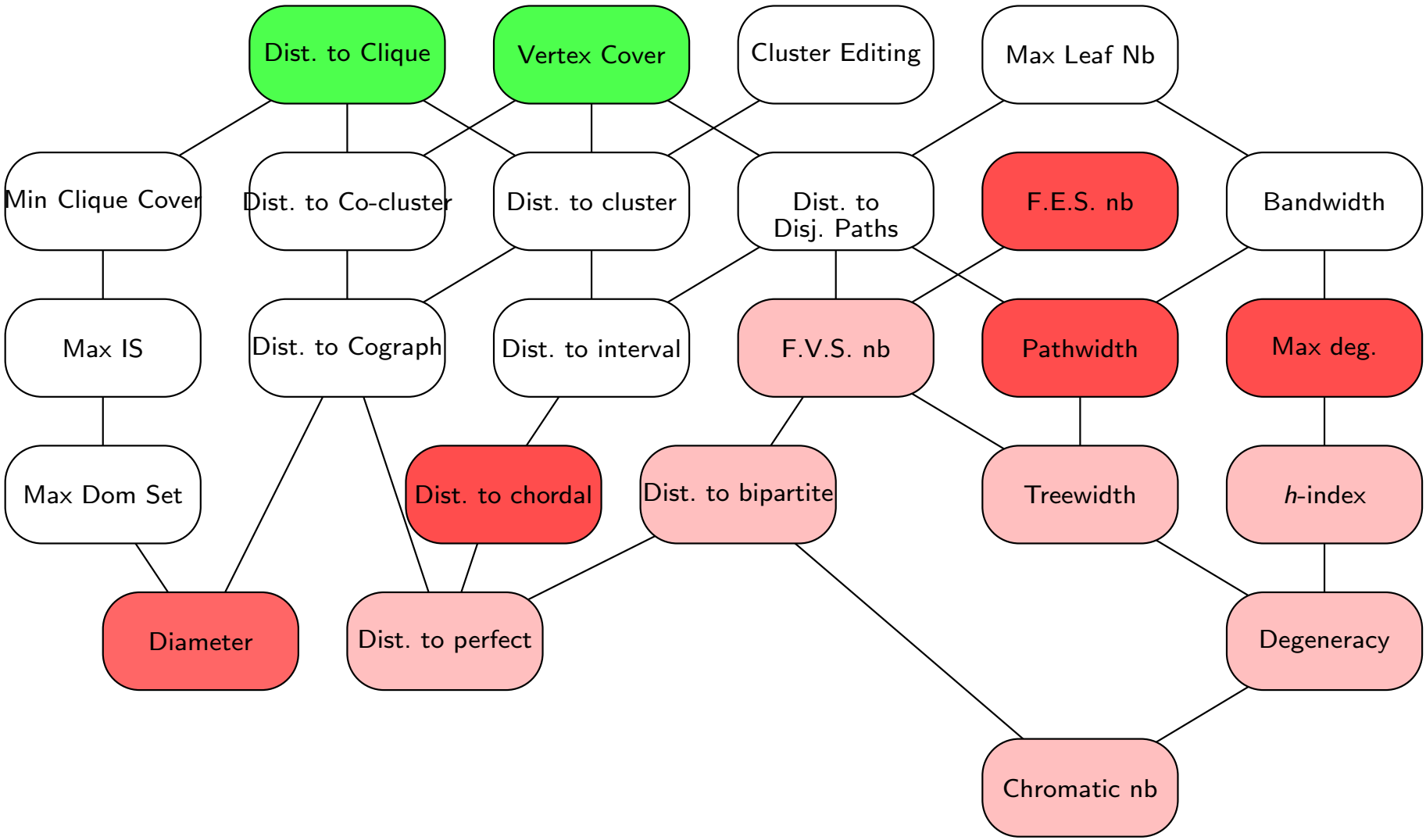
Perspectives

Perspectives



[Komusiewicz et Niedermeier, 2012]

Perspectives

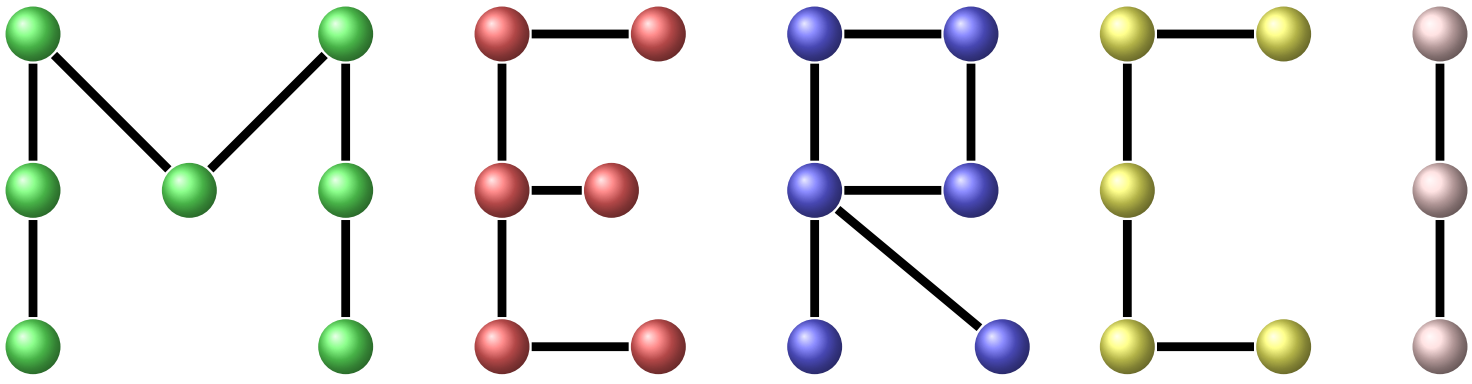


Perspectives

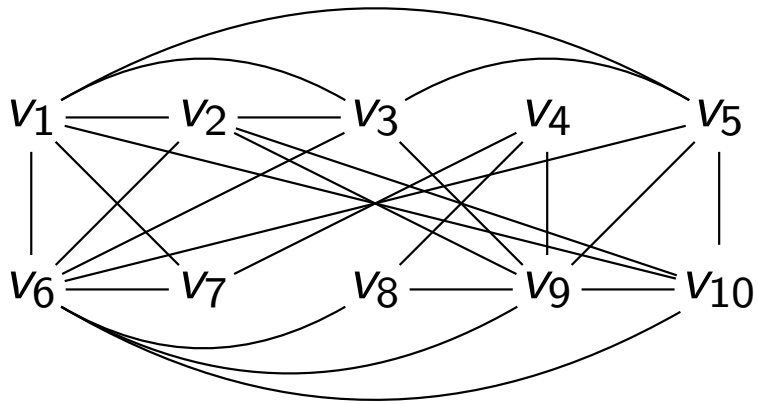
- Utilisation de paramètres structurels (adaptés aux données biologiques)
- Algorithme déterministe en espace polynomial ?

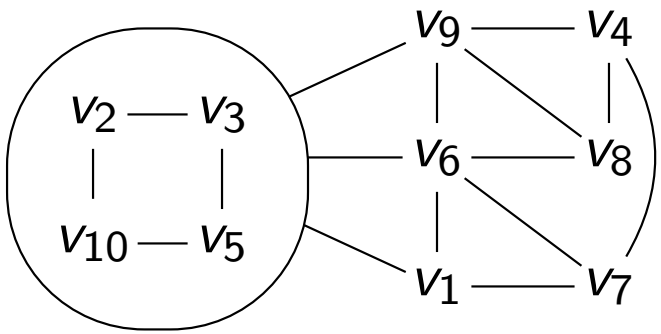
Perspectives

- Utilisation de paramètres structurels (adaptés aux données biologiques)
- Algorithme déterministe en espace polynomial ?
- Le bruit pousse à la recherche d'extensions plus souples :
  - Contraintes sur les couleurs.
  - Contraintes sur la connexité :
    - Simple connexité : FPT,
    - 2-connexité : W[1]-difficile,
    - Modules : FPT.
    - Alternatives ?

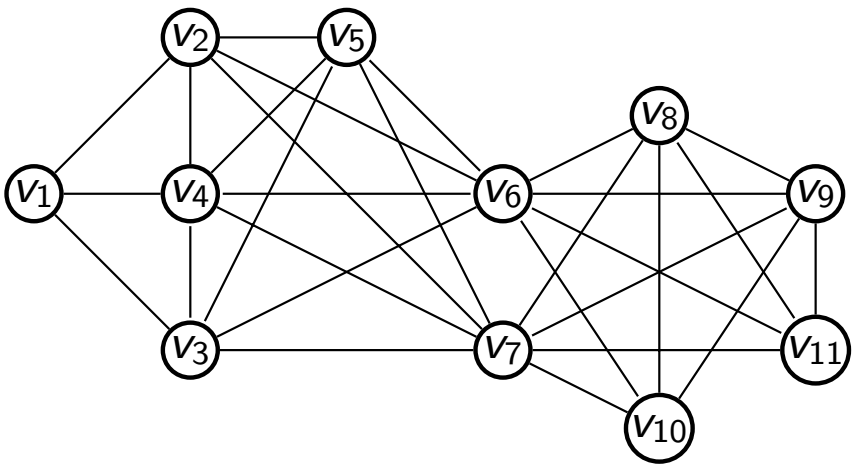

MERCI

Modules

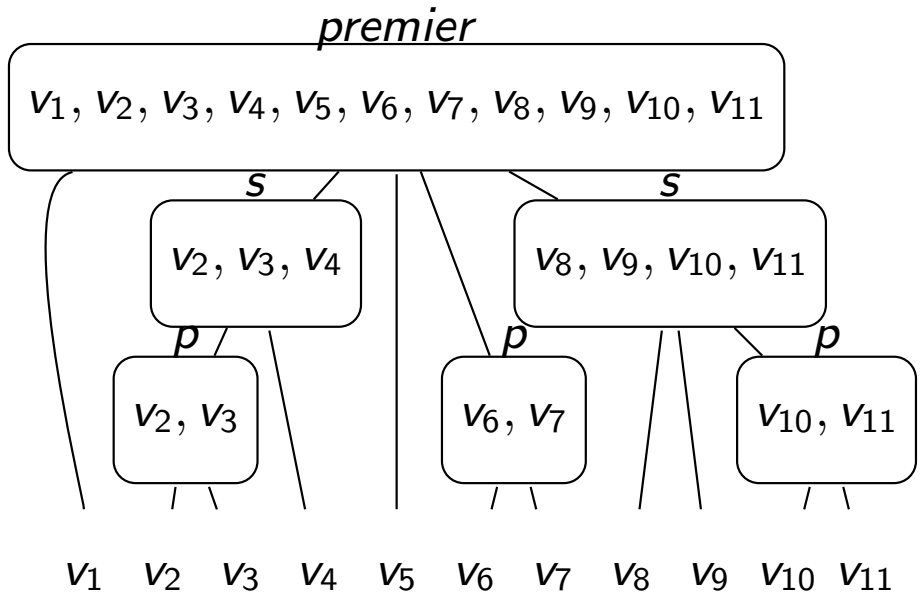




Modules


v_1
v_2
v_3
v_4
v_5
v_6
v_7
v_8
v_9
v_{10}
v_{11}

a)


premier
$v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}$
s
s
v_2, v_3, v_4
v_8, v_9, v_{10}, v_{11}
p
p
p
v_2, v_3
v_6, v_7
v_{10}, v_{11}
v_1 v_2 v_3 v_4 v_5 v_6 v_7 v_8 v_9 v_{10} v_{11}

b)