

---

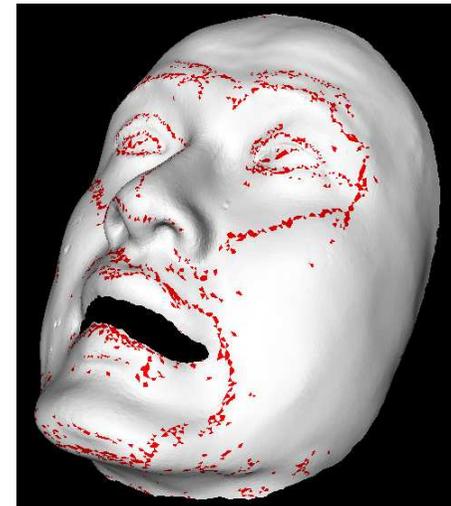
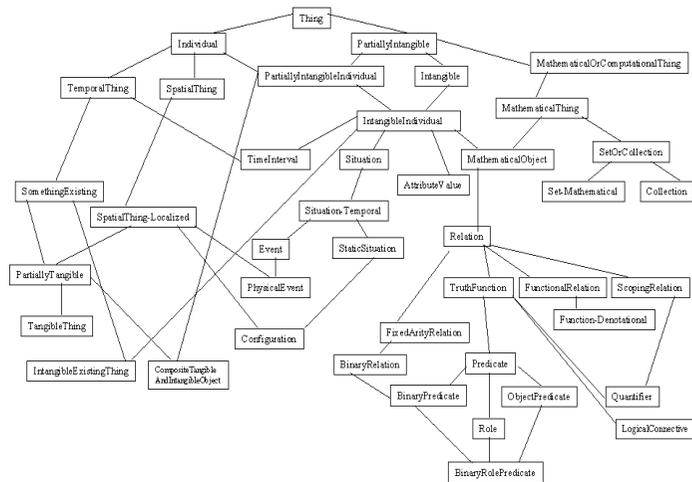
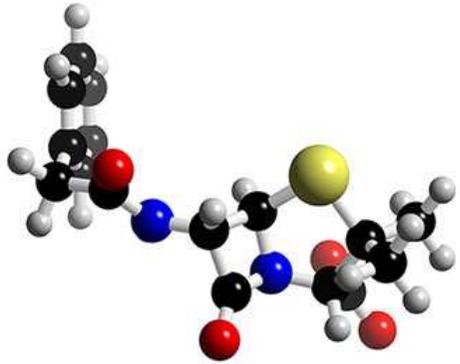
# Isomorphisme inexact de graphes par optimisation évolutionnaire

---

Thomas Bärecke



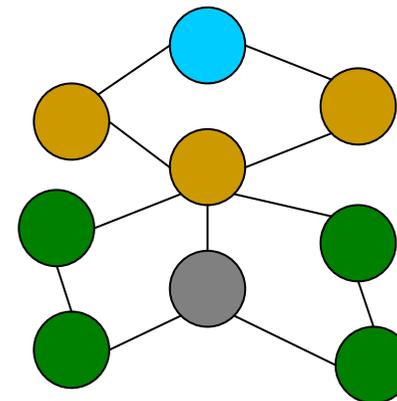
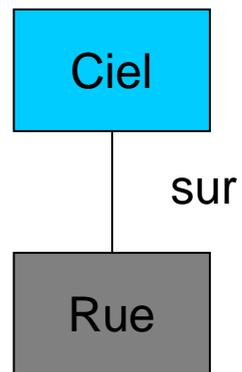
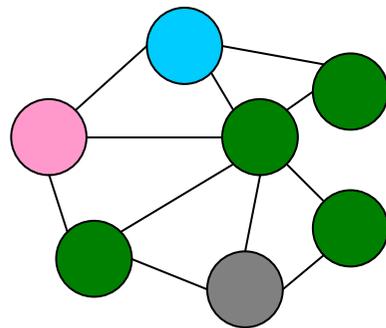
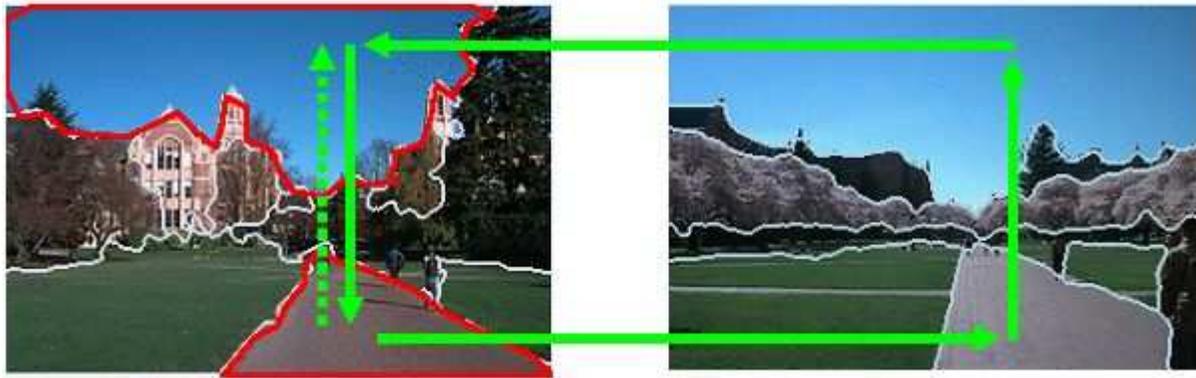
# Expressivité de la modélisation



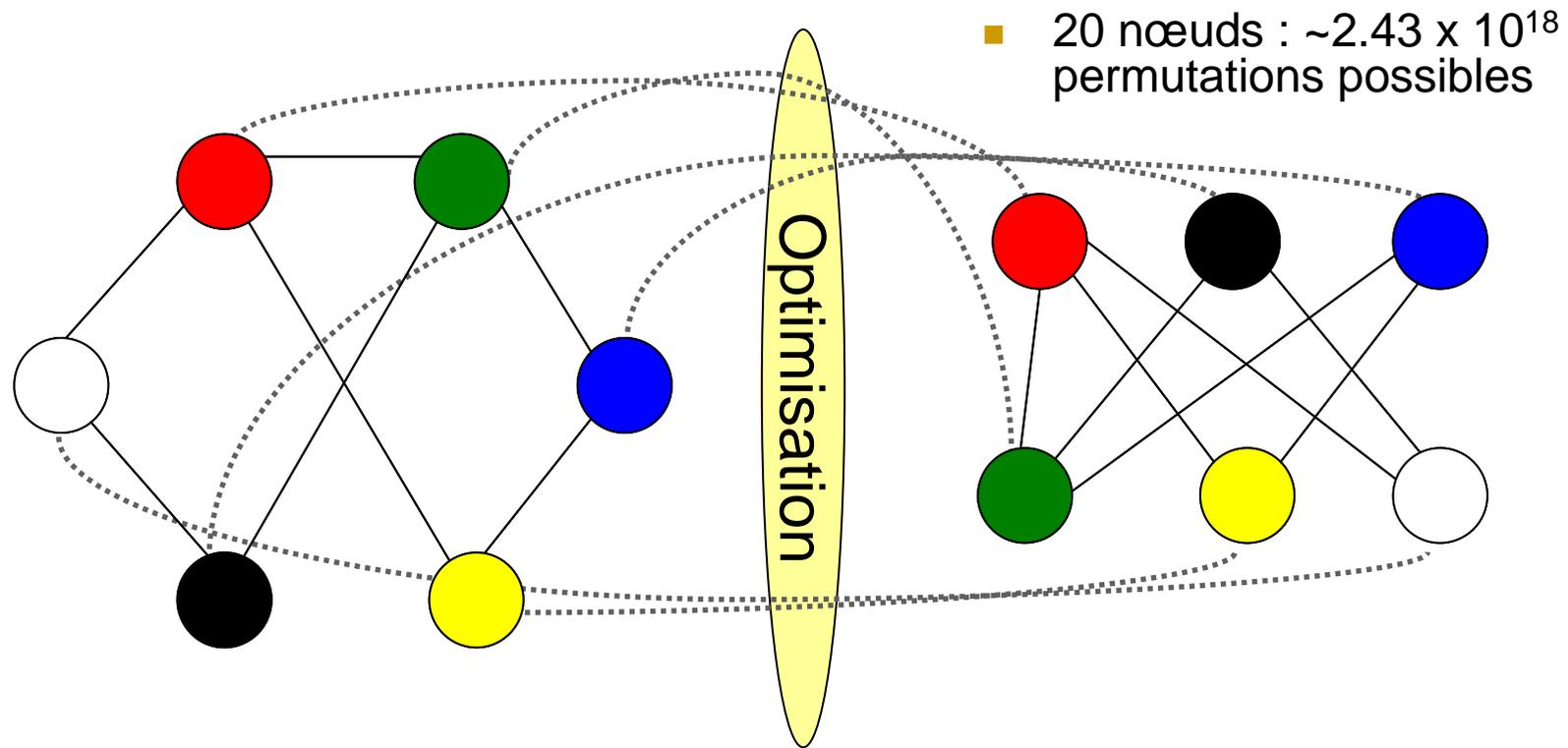
# Exemple d'application

## Recherche d'images

Omhover & Detyniecki ; Harchaoui & Bach



# Isomorphisme inexact de graphes



- Isomorphisme de sous-graphes
- Bruit  $\rightarrow$  isomorphisme inexact

# Isomorphisme inexact de sous-graphes

- Pas d'algorithme efficace **et** optimal
- Méthodes optimales (A\*)
  - Toujours **solution optimale**
  - Complexité (pire cas) **exponentielle**
  - Environ 15 nœuds max.
- Autres méthodes
  - Optimisation
  - Méthodes spectrales (Umeyama, 1998)
- Algorithmes évolutionnaires
  - Exploration parallèle d'un grand espace non continu
  - Pas d'exploitation parfaite
  - Critère d'arrêt configurable
    - Qualité de solution
    - Temps de calcul
  - **Bonne solution** dans un **temps raisonnable**
    - Suganthan, 2002 et Cross, 1997
    - Wang, 1997 et Torres-Velázquez, 2002
    - ACO - Sammoud, 2006

---

# Plan

- Introduction
- Cadre évolutionnaire
- Recherche locale
- Garantie du résultat optimal
- Applications réelles
- Conclusion et perspectives

# Formalisation de l'isomorphisme inexact

- Soit  $G$  et  $G'$  deux graphes :

$$G = (V, E, \alpha, \beta), \alpha : V \rightarrow L_V, \beta : E \rightarrow L_E$$
$$G' = (V', E', \alpha', \beta'), \alpha' : V' \rightarrow L_V, \beta' : E' \rightarrow L_E$$

- On cherche une fonction

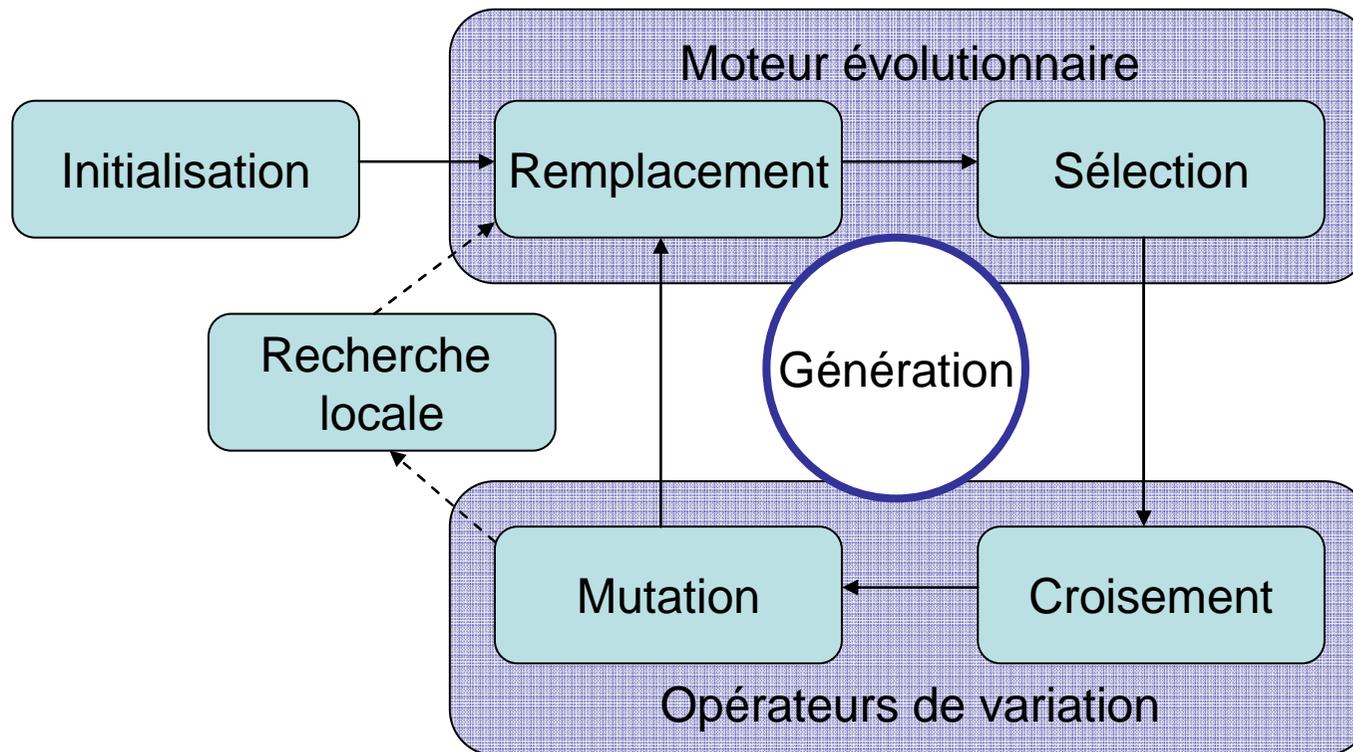
$$m : V \rightarrow V'$$

- qui minimise une mesure de distance

$$\arg \min_m \delta_m(G, G')$$

- Nous considérons la somme des distances d'attributs, de sommets et d'arêtes
- Optimisation combinatoire, espace factoriel

# Algorithme évolutionnaire



---

# Plan

- Introduction
- Cadre évolutionnaire
- Recherche locale
- Garantie du résultat optimal
- Applications réelles
- Conclusion et perspectives

---

# Éléments principaux

- Evaluation de la fonction de *fitness*
  - Distance de graphes sous l'appariement  $m$

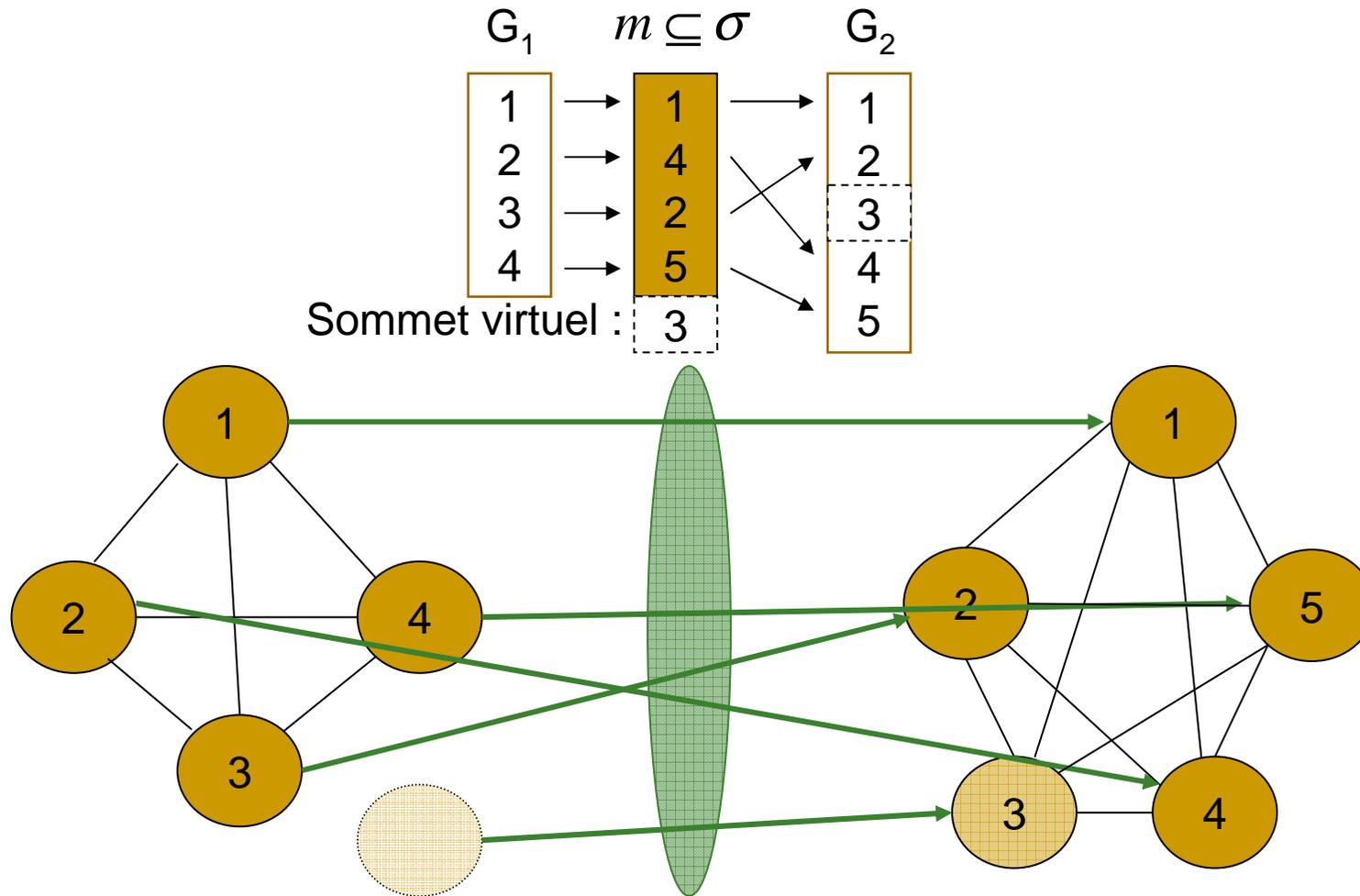
$$\delta_m(G, G')$$

- Complexité

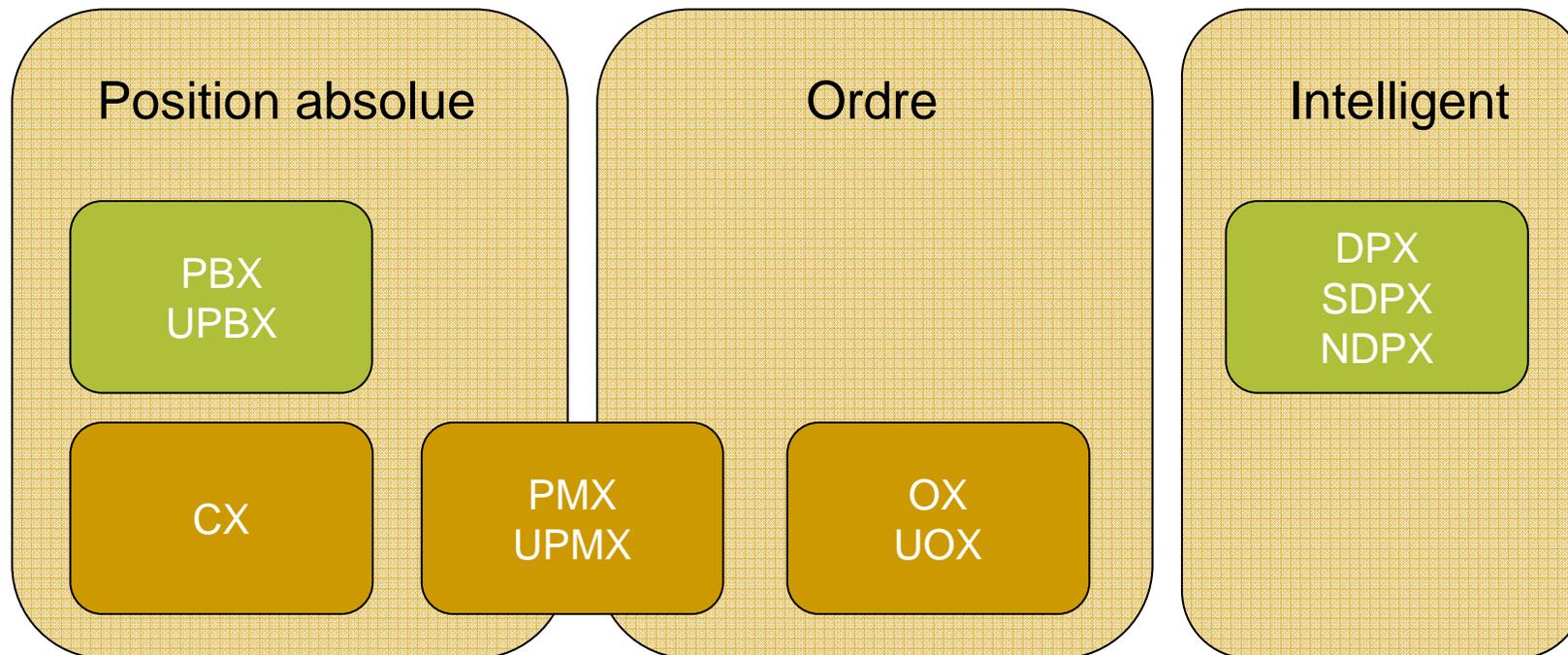
$$O(n^2)$$

- Codage
- Opérateurs de variation
  - Optimisation de paramètres
  - Comparaison

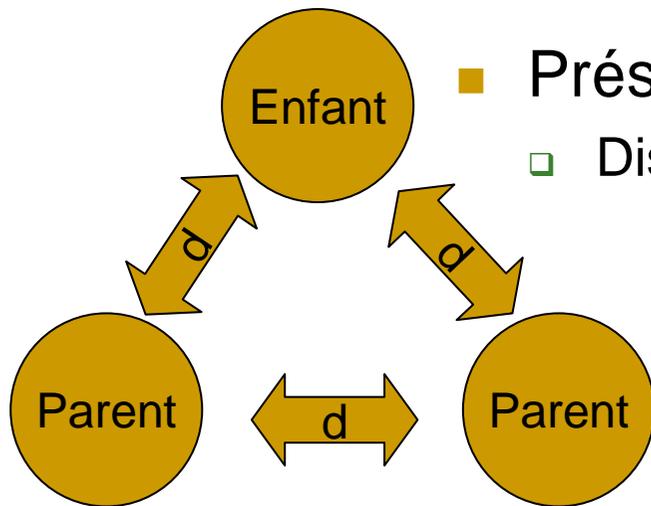
# Codage de permutation



# Opérateurs de croisement pour les permutations



# Croisements intelligents - DPX



- Préserve la distance entre les parents
  - Distance de Hamming dans l'espace génotypique

$$d_H(p_1, p_2) = d_H(e, p_{\{1;2\}})$$

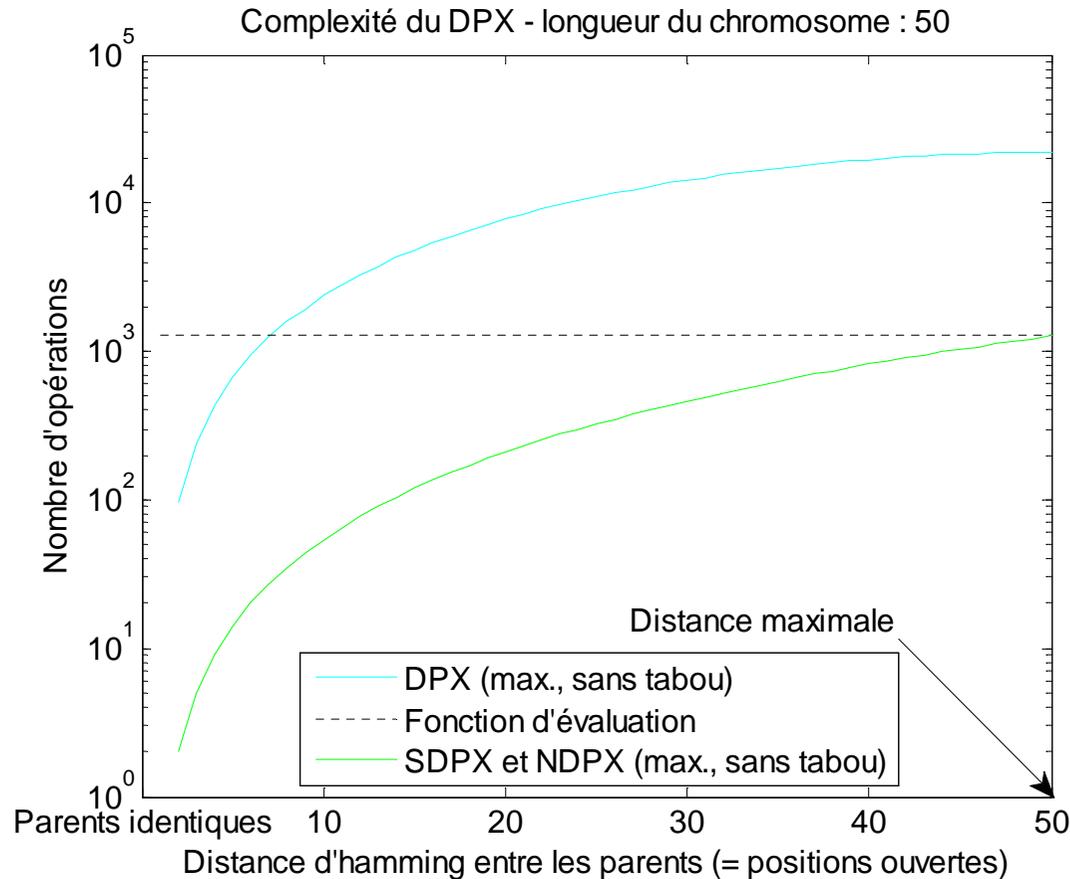
1. S'adapte aux différentes phases de l'algorithme
  - $d_H$  diminue
2. Remplissage glouton
  - Recherche d'un bon individu dans le rayon défini

---

# Variantes de DPX

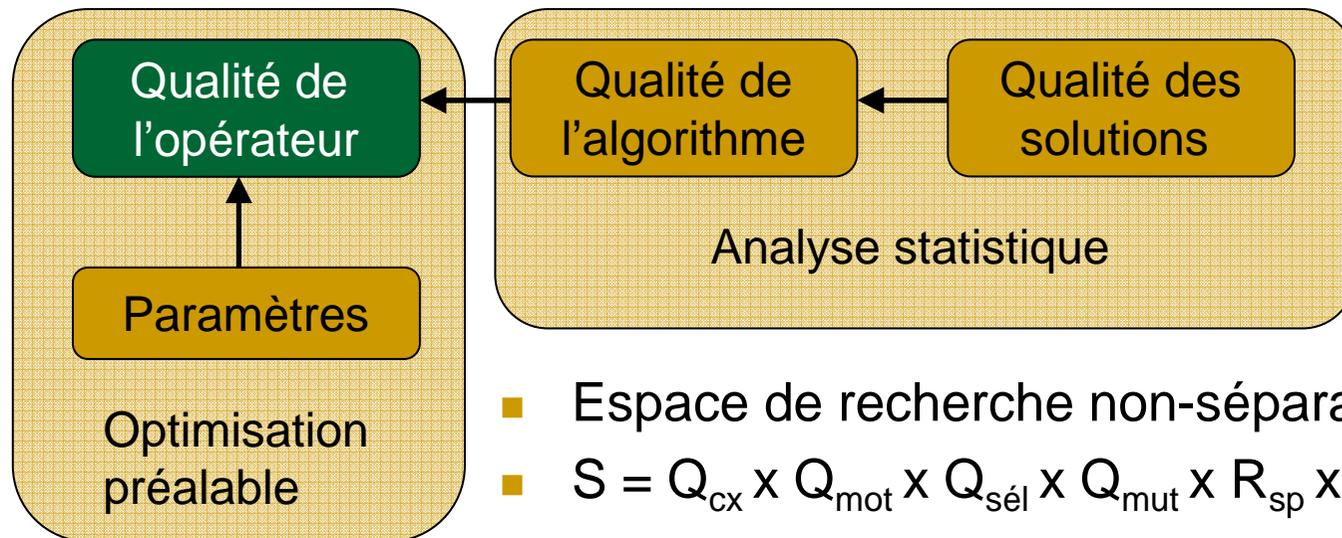
- Distance partielle à chaque étape
  - Complexité :  $O(na^2)$  ; a nombre de positions « ouvertes »
- Relâchement
  - NDPX (*Node DPX*)
    - Ignore les attributs d'arêtes
    - Complexité :  $O(a^2)$
  - SDPX (*Signature DPX*)
    - Nécessite une agrégation des attributs de sommets et des attributs d'arêtes incidentes
    - Complexité :  $O(a^2)$  + une fois le calcul d'agrégation

# DPX - complexité



- DPX est cher par rapport à la fonction d'évaluation
- Avec relâchement la complexité est dans le pire cas égale à celle de la fonction d'évaluation

# Comparaison des opérateurs

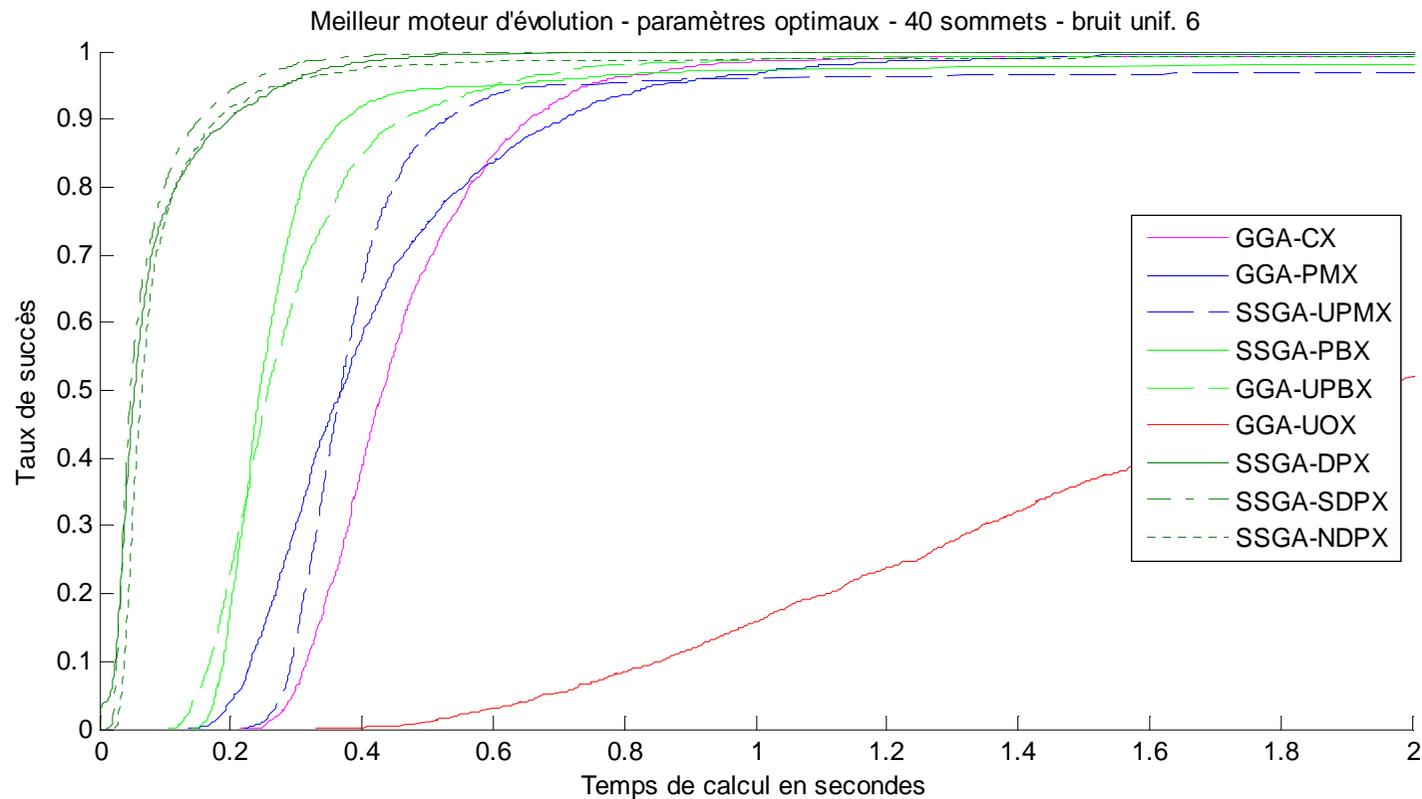


- Espace de recherche non-séparable
  - $S = Q_{cx} \times Q_{mot} \times Q_{sél} \times Q_{mut} \times R_{sp} \times R_{pm} \times R_{pc}$
- 
- Descente de gradient pour les paramètres numériques
    - Optimum local stable
  - 1500 itérations par combinaison de paramètres
  - x 2 moteurs x 9 opérateurs x 3 paramètres numériques

# Résultats par opérateur

Moteur	générationnel			<i>steady-state</i>		Sensibilité	Success performance
Opérateur	$s_p$	$p_c$	$p_m$	$s_p$	$p_m$		
SDPX	25	35%	30%	15	25%	$p_c$ Corrélation entre $p_c$ et $s_p$	0,074s
DPX	50	25%	25%	20	25%		0,086s (1)
NDPX	25	40%	30%	20	25%		0,105s
UPBX	90	99%	20%	400	6%	$p_m$	0,298s (2)
PBX	120	96%	30%	300	30%		0,315s
UPMX	90	80%	12%	400	6%	$s_p$ et $p_c$	0,429s
PMX	120	99%	40%	1000	70%	$p_m$	0,430s (3)
CX	400	99%	60%	700	60%	$s_p$ - taille des graphes	0,472s
UOX	10	60%	14%	400	7%	Tous les paramètres	2,372s (4)

# Comparaison globale d'opérateurs – meilleur moteur

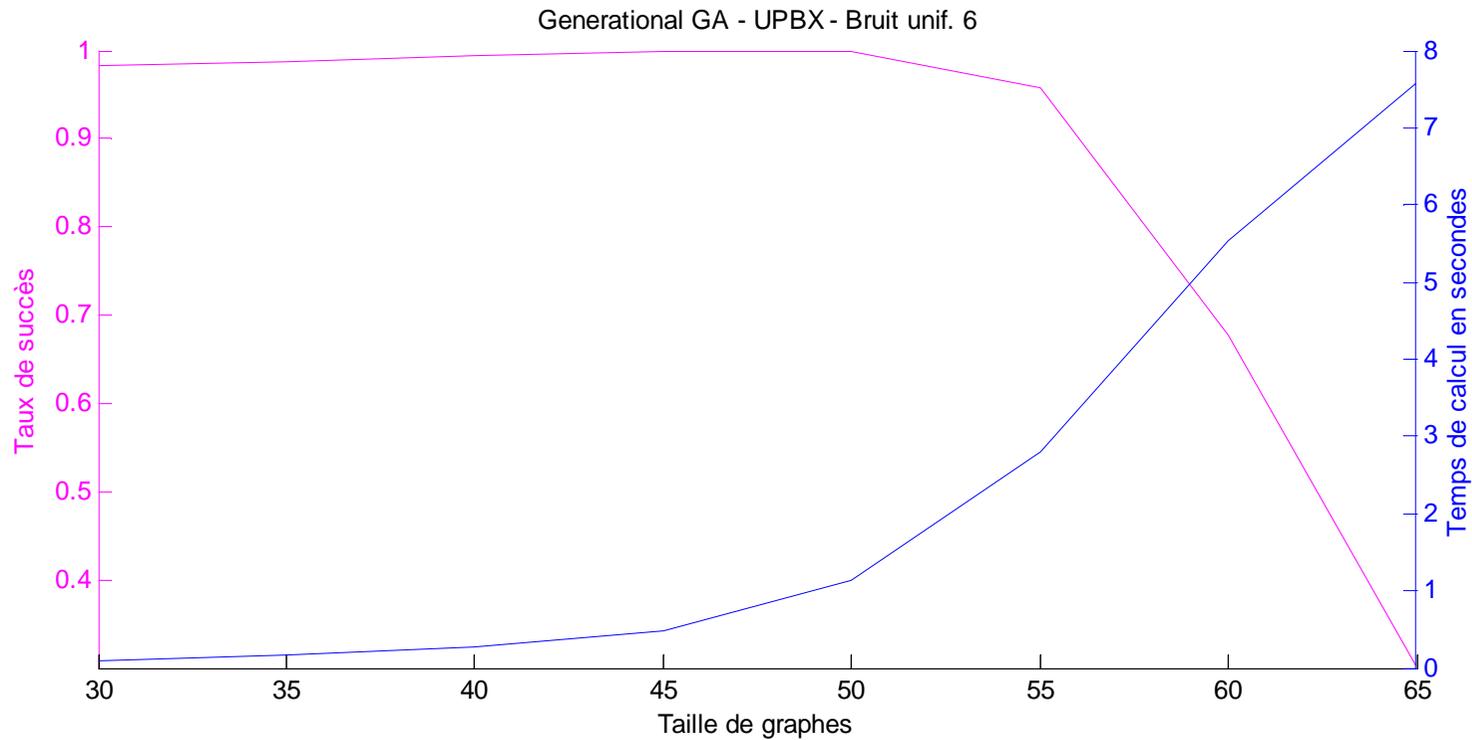


---

# Plan

- Introduction
- Cadre évolutionnaire
- Recherche locale
- Garantie du résultat optimal
- Applications réelles
- Conclusion et perspectives

# Limite d'UPBX



- Perte de précision pour des graphes plus grands

# Recherche locale

- Voisinage  $N$  d'un individu

$$N(\sigma) = \{\sigma' = M(\sigma)\}$$

- Opérateur *two-opt*
- Évaluation d'un élément du voisinage
  - prend en compte uniquement les affectations modifiées

$$O(n) \text{ au lieu de } O(n^2)$$

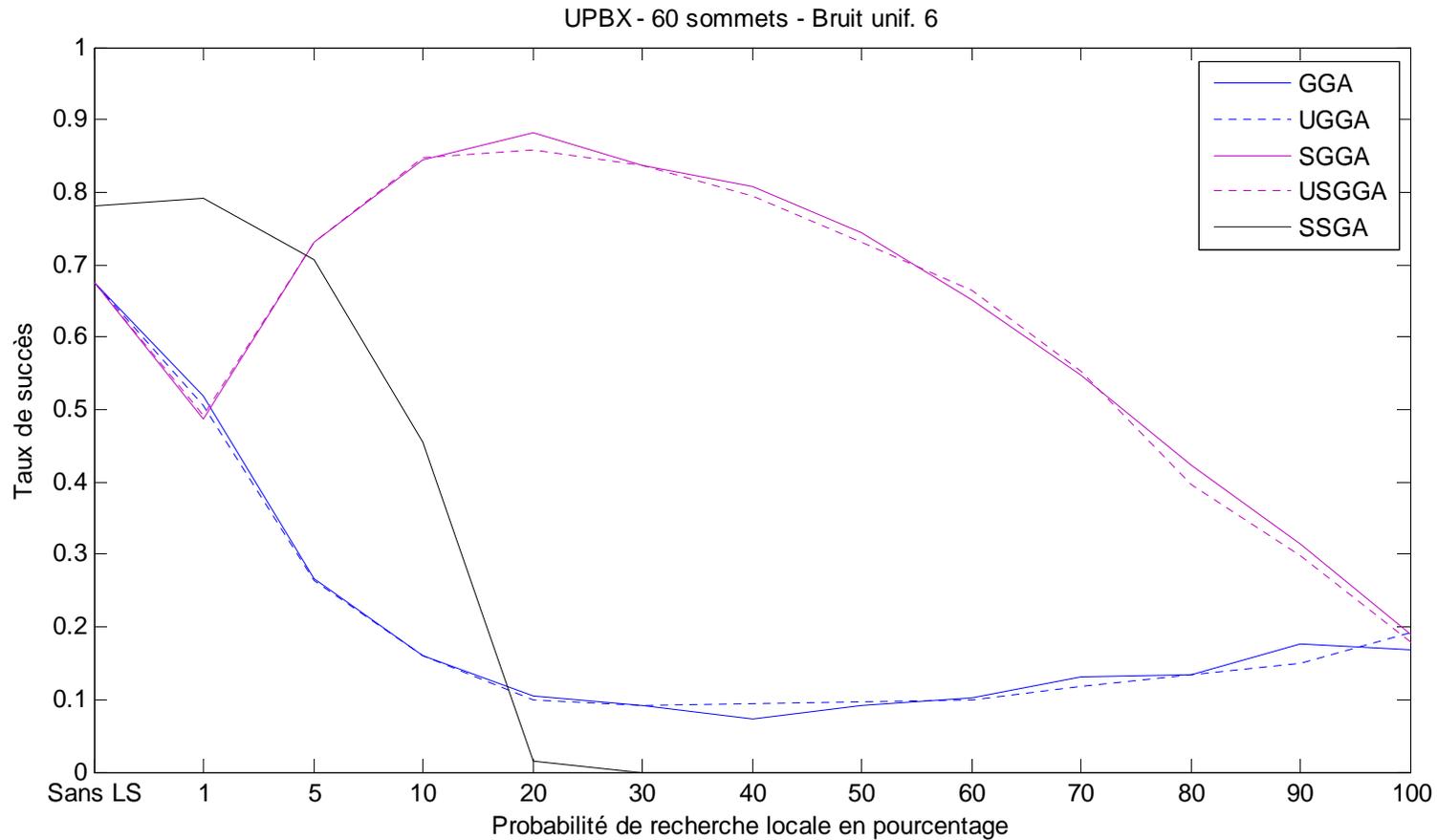
- Évaluation de l'espace de voisinage  $O(n^3)$

---

# Stratégies d'intégration

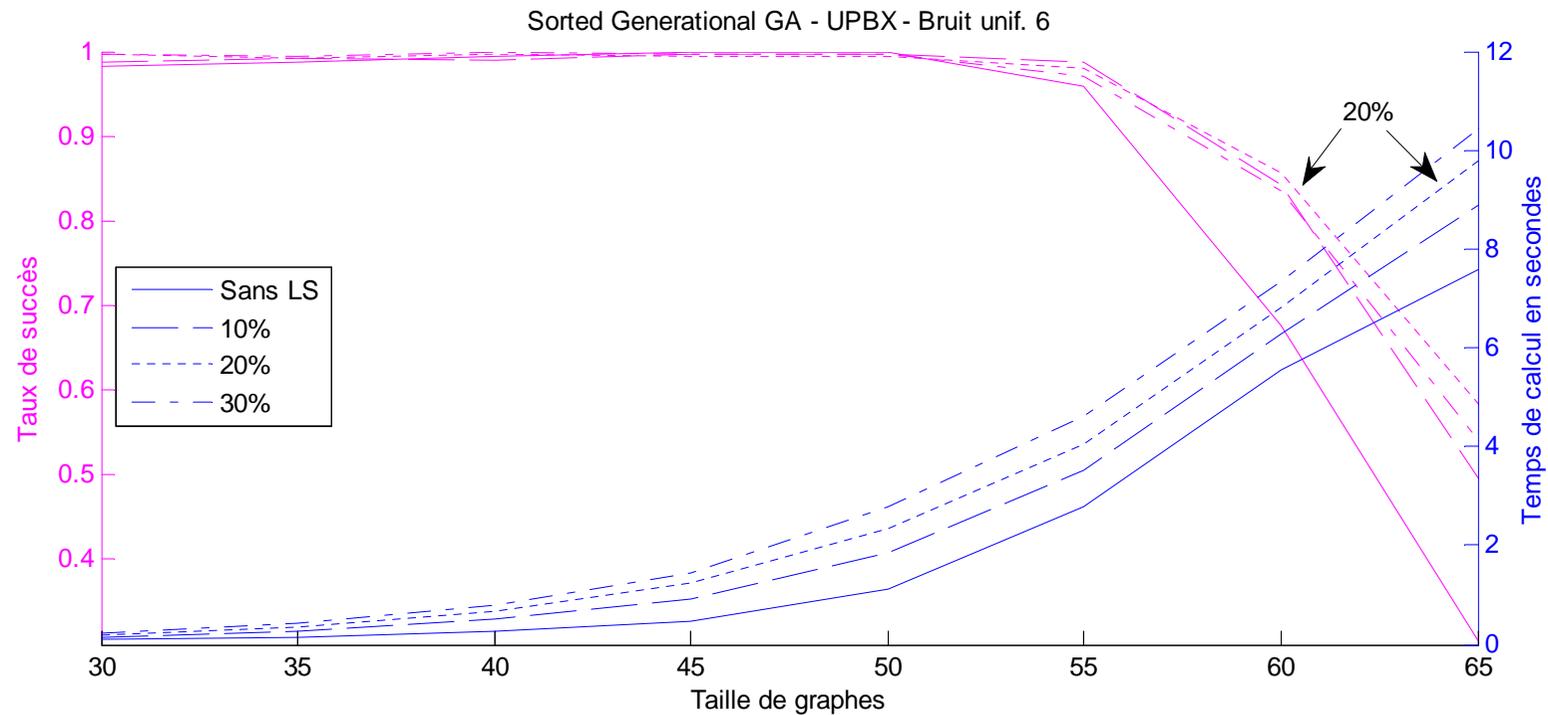
- Répartition exploration – exploitation
- G(enerational)GA et S(teady)S(tate)GA
- Chaque individu a la même probabilité
- U(nique)GGA
  - Exclut les copies
- S(orted)GGA
  - Applique la recherche locale aux meilleurs individus uniquement
- USGGA
  - Combine les deux précédents

# Comparaison des stratégies (GGA-UPBX)



- Versions triées >> non-triées
- Probabilités généralement faibles

# Limite UPBX avec recherche locale



- Meilleure précision
- Convergence légèrement ralentie
- La recherche locale permet un meilleur passage à l'échelle
  - appliquée à une faible partie de la population
  - appliquée aux meilleurs individus

# Plan

- Introduction
- Cadre évolutionnaire
- Recherche locale
- Garantie du résultat optimal
- Applications réelles
- Conclusion et perspectives

---

# Garantie du résultat optimal

- Méthode de recherche exhaustive  $A^*$ 
  - Seuil d'acceptation
  - Coût de calcul élevé
- Cascade
  - Elagage de l'espace de recherche
  - Grâce à une borne stricte fourni par l'approche évolutionnaire

# Recherche exhaustive – $A^*$

Appariement vide



Une paire de nœuds



Deux paires de nœuds

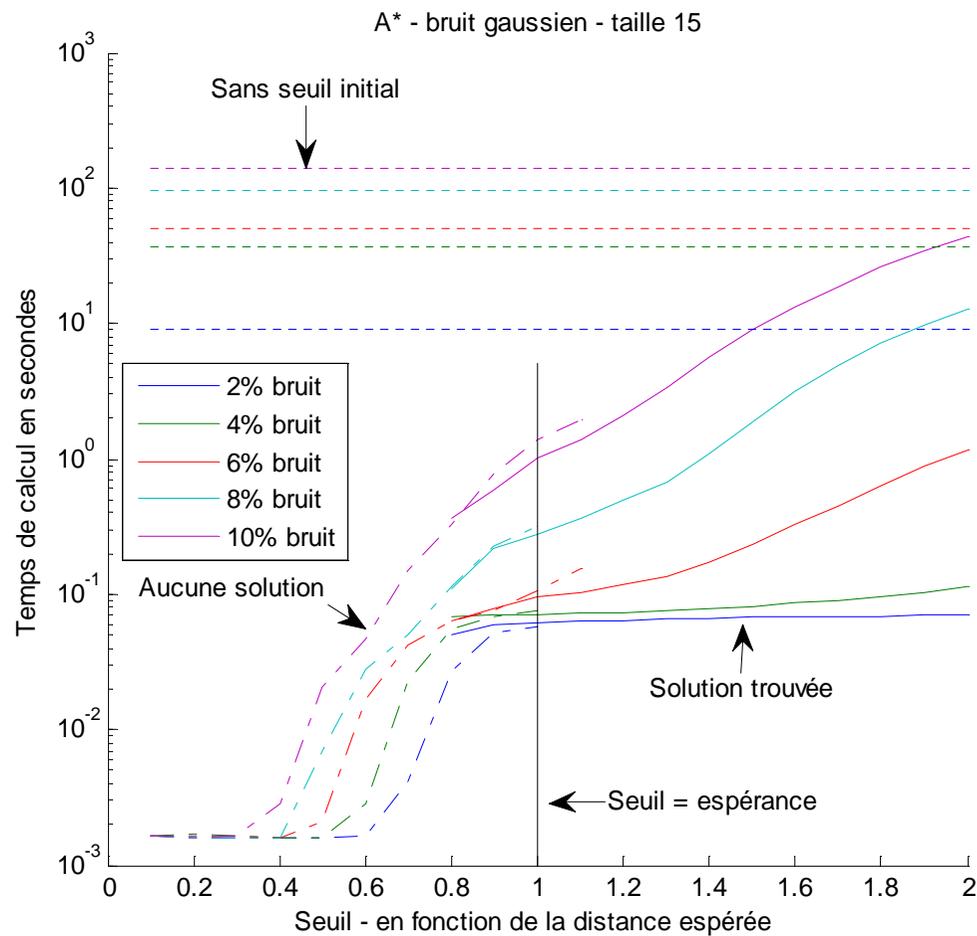


Appariement complet



L'estimation d'une distance maximale admissible permet d'élaguer l'espace de recherche.

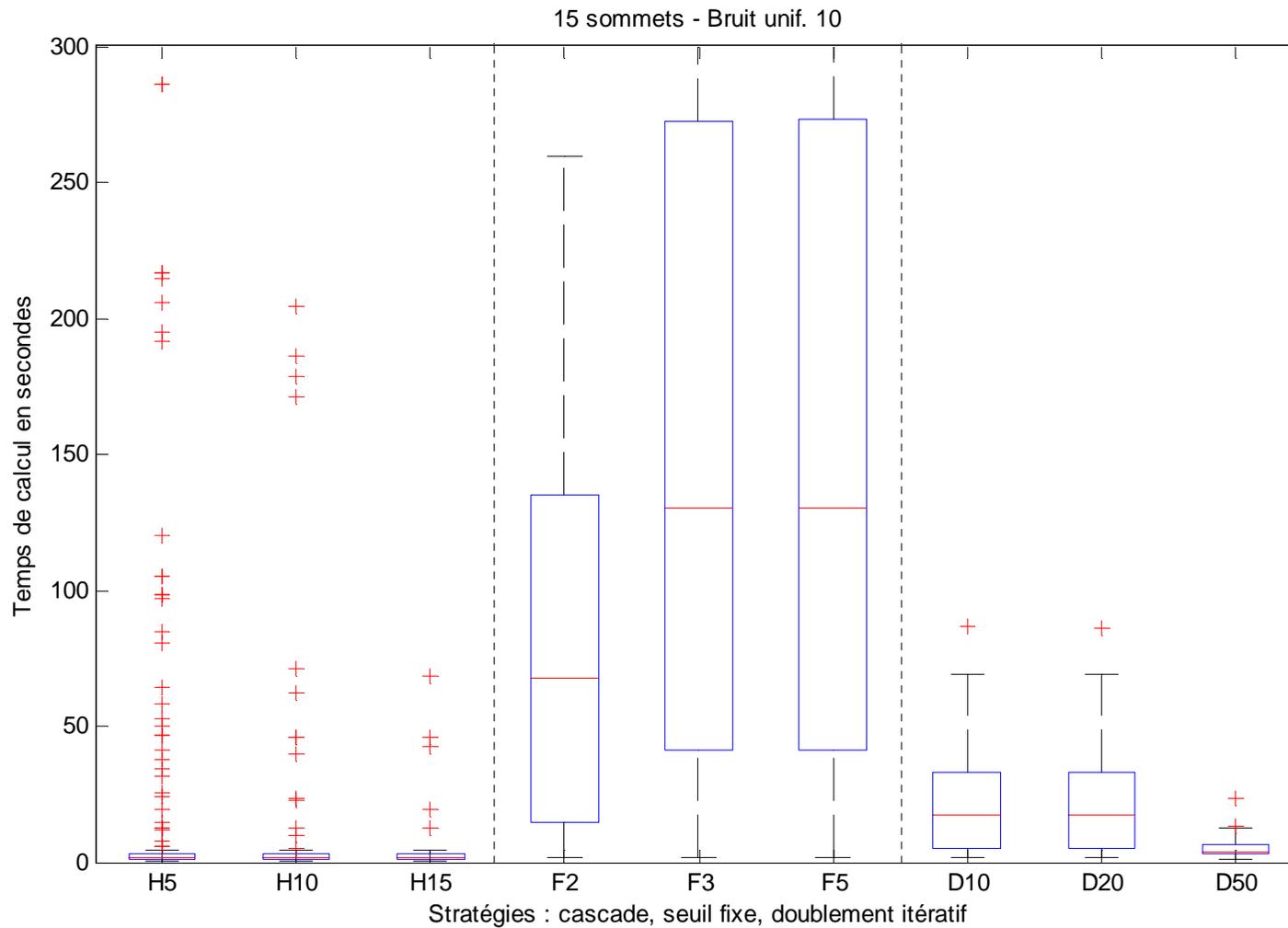
# Influence du seuil



## Variante itérative

- Seuil initial faible
- Croissance exponentielle

# Comparaison des stratégies



---

# Plan

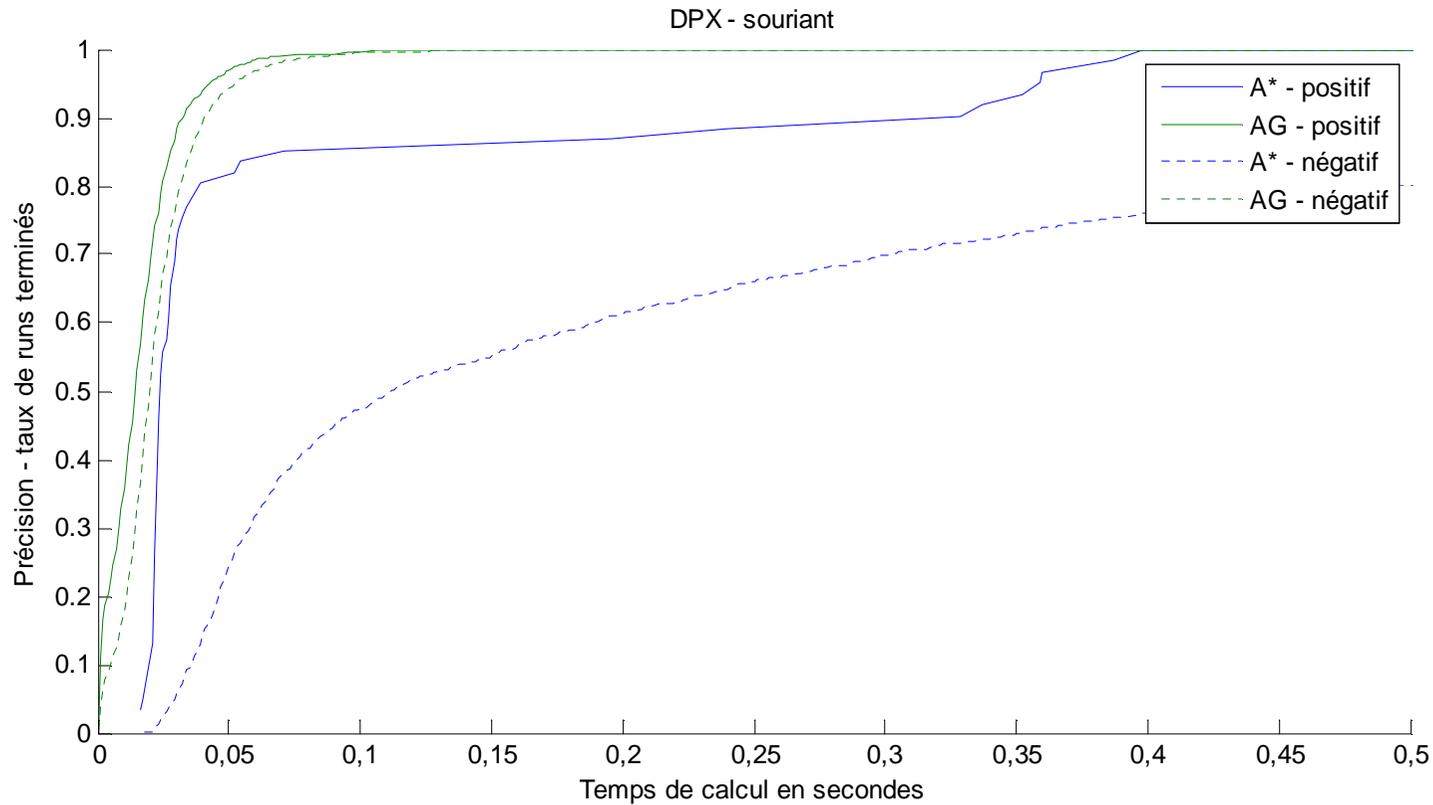
- Introduction
- Cadre évolutionnaire
- Recherche locale
- Garantie du résultat optimal
- **Applications réelles**
- Conclusion et perspectives

---

# Applications réelles

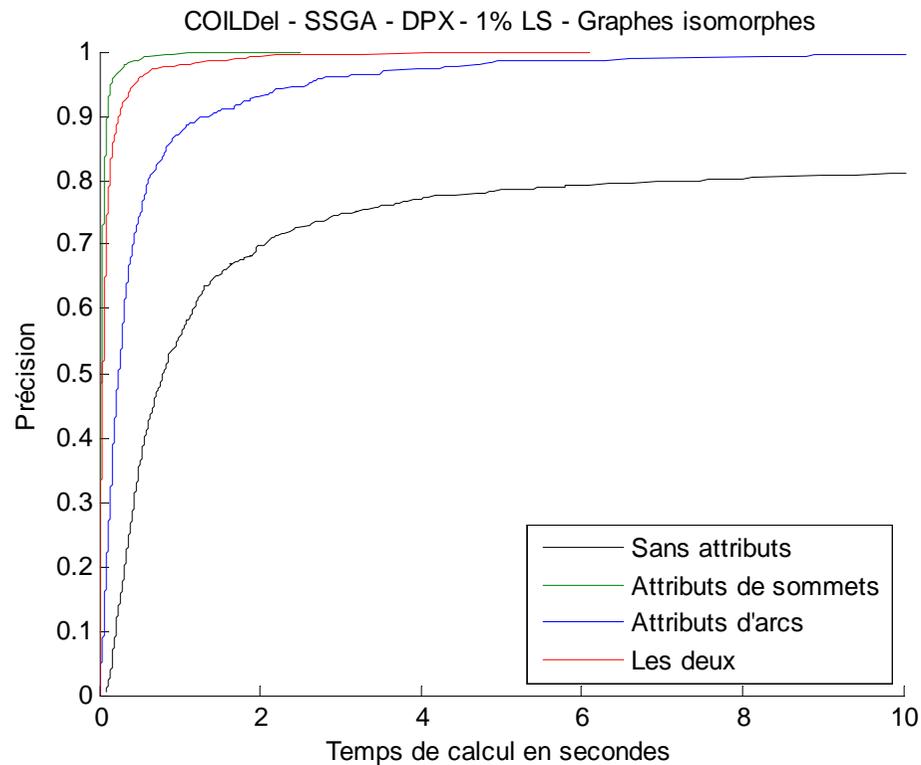
- Images 3D – Berretti
  - 61 visages, 3 prises de vue (modèle 3D, photo frontale neutre/sourire)
  - Graphes petits et étiquetés
  - Isomorphisme inexact - variance d'attributs
- COIL-Del – Riesen
  - Triangulation de Delaunay des points d'intérêts
  - Graphes plus grands sans attribut
  - Isomorphisme inexact – fortes perturbations structurelles
- Molécules – *National Cancer Institute*
  - Isomorphisme exact

# Images 3D



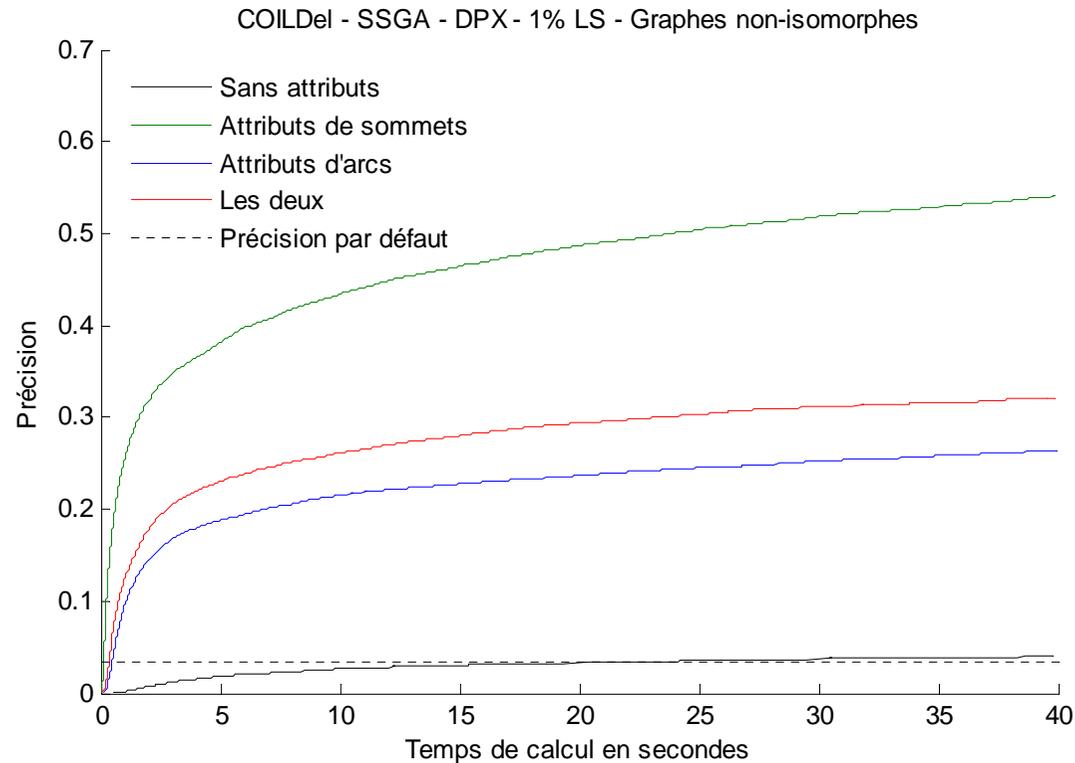
- Expression faciale souriante – fortes perturbations
- Meilleure performance que la référence

# COIL – isomorphisme exact



- Meilleures performances avec l'utilisation d'informations complémentaires

# COIL – graphes non isomorphes



- Fortes perturbations structurelles
- Performances améliorables

---

# Conclusion

- Codage générique pour graphes et sous-graphes
- Bonne robustesse aux perturbations d'attributs
- Dépassement de la limite de taille de l'état de l'art
  - Meilleurs opérateurs de variation
  - Recherche locale appliquée à l'« élite »
- Etude des stratégies de parallélisation
- Garantie de l'optimalité par cascade

---

# Perspectives à court terme

- Comparaison à d'autres méthodes
  - En particulier : noyaux, colonies de fourmis
- Augmentation du rayon d'application
  - Opérateurs intelligents
  - Recherche locale
  - Mutation
- Réglage de paramètres *versus* contrôle automatique de paramètres
- Modèle parallèle en îlots

---

# Perspectives à long terme

- Etude d'autres mesures de distance
  - Robustesse au bruit structurel
- Développement d'applications d'isomorphisme à grande échelle
- Appariements multivoques
  - Approche co-évolutionnaire

