

Équipe algorithmique du LabInfo IGM, UMR CNRS 8049

5 février 2009



- Responsable : Marie-Pierre Béal (jusqu'en 2008)
Sylvain Lombardy
- 22 membres permanents :
 - 13 maîtres de conférences (-2 +5)
 - 4 professeurs (-3 +1)
 - 3 CR CNRS (-1 +4)
 - 2 professeurs émérites (+2)
- 8 doctorants, 2 post-doctorants
- 14 invités (2004-2008) Europe, Canada, USA, Japon,...

Publications et logiciels

- \simeq 85 articles en revues
- \simeq 170 conférences avec actes
- 5 directions d'ouvrages collectifs
- 23 chapitres d'ouvrages scientifiques
- 3 ouvrages ou chapitres de vulgarisation
- 14 thèses de doctorat
- 3 habilitations à diriger des recherches
- 15 logiciels

Ouvrages

- Algorithms on Strings, de M. Crochemore, C. Hancart et T. Lecroq
- Codes and Automata, de J. Berstel, D. Perrin et Ch. Reutenauer
- Combinatorics on Words: Christoffel Words and Repetitions in Words, de J. Berstel, A. Lauve, Ch. Reutenauer et F.V. Saliola
- Combinatorics of Genome Rearrangements (à paraître), de G. Fertin, A. Labarre, I. Rusu, É. Tannier et S. Vialette
- Transmissions et Réseaux, de S. Lohier et D. Présent

Chapitres

- Chapitre dans "Handbook of Information Theory and Coding", par M.-P. Béal, J. Berstel, B. Marcus, D. Perrin, Ch. Reutenauer et P. Siegel
- Chapitre dans "Encyclopédie de l'informatique et des systèmes d'information", par J. Berstel et L. Boasson
- Chapitres dans "Les réseaux sans fils émergents" et "Contrôle dans les réseaux IP", par S. Lohier

Édition

- Œuvres de M.-P. Schützenberger, par J. Berstel, D. Perrin et A. Lascoux
- Projet Lothaire: "Applied Combinatorics on Words", par J. Berstel et D. Perrin

Intervention dans des cursus d'enseignement

- Licence Mathématique Informatique
- École Ingénieurs 2000, filière Informatique et Réseaux
- Master informatique
 - filière Algorithmique et Bio-informatique
 - filière Réseau
- Professeur à King's College, Londres

Participation aux comités de rédaction

- Theoretical Computer Science
- Rairo Theoretical Informatics and Applications
- Computational Biology and Chemistry
- IEEE Communications Surveys and Tutorials
- Trends in Applied Science Research journal
- Software Engineering Journal
- Comités de programmme : ICALP, STACS, FOSSACS, CPM, SPIRE, ACM SAC PL, RECOMB, IEEE Globecom, . . .
- Comités de pilotages: DLT, CAI, CPM, WORDS

- ALEA 2007
- Workshop on Symbolic Dynamics and Coding 2007
- Journées conjointes GDR BiM - GDR IM 2007
- RESCOM 2008 (Réseaux et Communications)
- Journées doctorales informatiques et réseaux 2007
- Frac hiver 2009

- Bilatéraux (Espagne, Italie, Uruguay, Taiwan, Canada, Slovénie)
- ANR
 - Gamma : Génération aléatoire
 - Brasero : Comparaison de structures d'ARN
 - Sycomore : Systèmes complexes et modèles de calcul
 - Stamp : Modélisation et analyse spatiale
- ACI Masse de données - ACI Indexation de texte
- AS Nouveaux modèles et algo. de graphes pour la biologie
- Projet européen AutoMathA
- GDR IM, groupes SDA2 et COMATEGE
GDR Bioinformatique Moléculaire
- Contrat EsaLab : Détection de contrefaçon logicielle
- Thèses CIFRE : EADS-Astrium, Exalead

- Algorithmique sur les mots, combinatoire des mots
- Automates, codage et dynamique symbolique
- Graphes infinis
- Analyse d'algorithmes
- Algorithmes pour la bio-informatique
- Langage et programmation
- Protocoles et réseaux

Algorithmique sur les mots et automates

- Localisation de motifs bi-dimensionnels
- Structures pour l'indexation
- Combinatoire liée à la compression
- Traitement de données musicales
- Recherches de similarités dans des codes sources
- Mots sturmiens et épisturmiens, arbres sturmiens
- Analyse en moyenne
- Automates sur les ordres
- Automates pondérés, séries rationnelles
- **Logiciels** REGAL, Vaucanson
- **Groupe de Travail** Automates

Codage et dynamique symbolique

- Dynamique symbolique et codage
- Numération non standard
- Traces d'automates cellulaires
- Automates de sable
- Complexité de Kolmogorov

Graphes infinis

- Algorithmique des structures infinies
- Lien entre graphes infinis et théorie des langages
- Vérification automatique

Algorithmes pour la bio-informatique

- Alignement
- Bases de motifs à jokers
- Inférence de motifs structurés
- Algorithmique et combinatoire des structures d'ARN
- Réarrangement génomiques
- Invariance des réseaux d'interactions
- **Logiciels** MiGAL, Ed'Nimbus
- **Groupe de Travail** AlgoB

Langages et Programmation

- Multi-méthodes Java
- Machine virtuelle paramétrable
- Génération de parseur
- **Logiciels** Tatoo, Corosol, Saburo,...

Algorithmique et programmation temps-réel

- Traitement des tâches apériodiques
- Bornes sur le temps d'exécution
- **Logiciels** RT Simulator, MEDL
- **Groupe de Travail** AlgoTR

Protocoles et réseaux

- Algorithmes de routage et QoS dans les réseaux ad hoc
 - Réseaux de capteurs multimédia sans fil
 - Réseaux MIMO
- Architecture inter-couches – QoS dans les WLANs
- Hiérarchie des réseaux sans fils et passage inter-systèmes
- Logiciel Qolyester
- Groupe de Travail PASNet

Recherche fondamentale et appliquée

- Langages et automates
- Algorithmique du texte
- Prog. et algo. temps-réel
- Codage
- Bio-informatique
- Protocoles et réseaux

Algorithmes pour la bio-informatique

- Approximation et complexité paramétrée
- Recherche de motifs dans les réseaux biologiques
- Comparaison de séquences avec duplication

Analyse en moyenne

- Complexité en moyenne sur les automates
- Bornes asymptotiques
- Comparaison des modèles probabilistes

Combinatoire des mots - langages

- Distribution des répétitions
- Étude des mots riches en palindromes
- Structure des mélanges de langages

Graphes infinis

- Étude des graphes d'ordres supérieurs
- Application aux langages des graphes déterministes
- Élaboration d'un outil pour la vérification

Automates pondérés

- Représentations déterministes de séries
- Étude combinatoire des automates à groupe

Codage et dynamique symbolique

- Complexité du "road colouring"
- Étude des codes bifixes
- Traces des sous-shifts non sofixes

Langage et Programmation

- Développement de Tadoo
- Algorithmes de comparaison de code guidés par grammaires
- Java temps réel

Protocoles et Réseaux

- Routage pour les réseaux de capteurs
- QoS pour les réseaux MIMO ad hoc

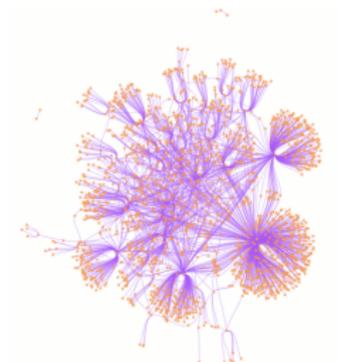
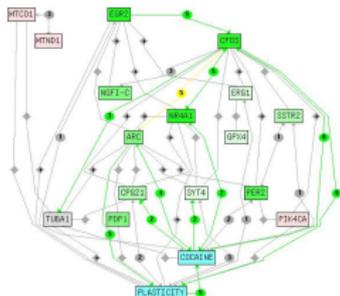
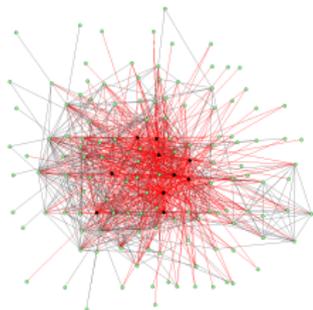
Recherche de Motifs Connexes dans les Réseaux Biologiques

5 février 2009



Réseaux biologiques

- L'étude des propriétés des réseaux biologiques est un sujet en pleine expansion.
- On entend par réseaux biologiques une grande variété de structures :
 - réseaux métaboliques,
 - réseaux d'interactions entre protéines,
 - réseaux de régulation,
 - ...



Requêtes dans les réseaux biologiques

Definition (REQUÊTE (problème générique))

Données:

- Un réseau biologique G , et
- un motif H .

But: Identifier dans le réseau G des “*motifs similaires*” à H , la similarité pouvant être définie en termes

- de séquences, et/ou
- de patterns d'interaction.

Limitations

La plupart des approches existantes nécessite la connaissance précise du pattern d'interaction de la requête:

Definition (MOTIF CONNEXE)

Données:

- Un ensemble de couleurs C ,
- un motif \mathcal{M} de taille k sur C (i.e., un multiensemble \mathcal{M} de support C),
- un graphe $G = (V, E)$ de taille n , et
- une coloration des sommets du graphe G $\lambda : V \rightarrow C$.

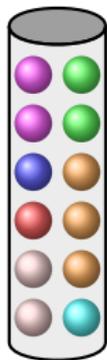
But: Trouver une occurrence connexe de \mathcal{M} dans G , i.e., un sous-ensemble $V' \subseteq V$ tel que

- $\lambda(V') = \mathcal{M}$, et
- $G[V']$ est une sous-graphe induit connexe.

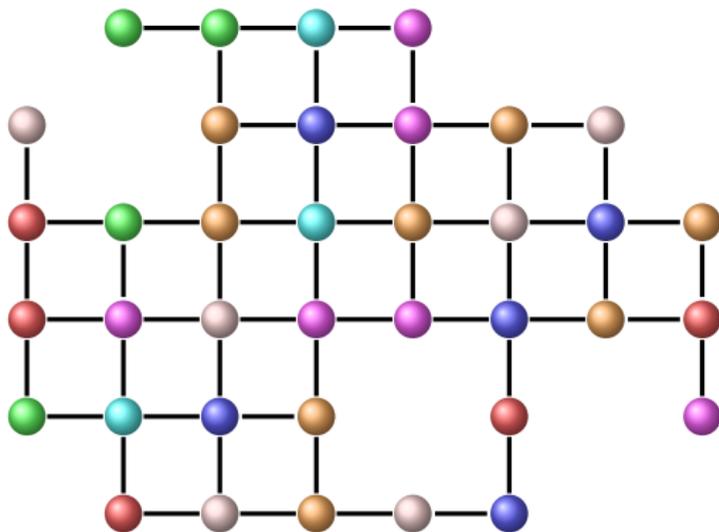
Remarks

Le motif \mathcal{M} est dit **simple** s'il est un ensemble.

MOTIF CONNEXE

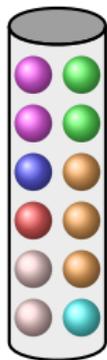


\mathcal{M}

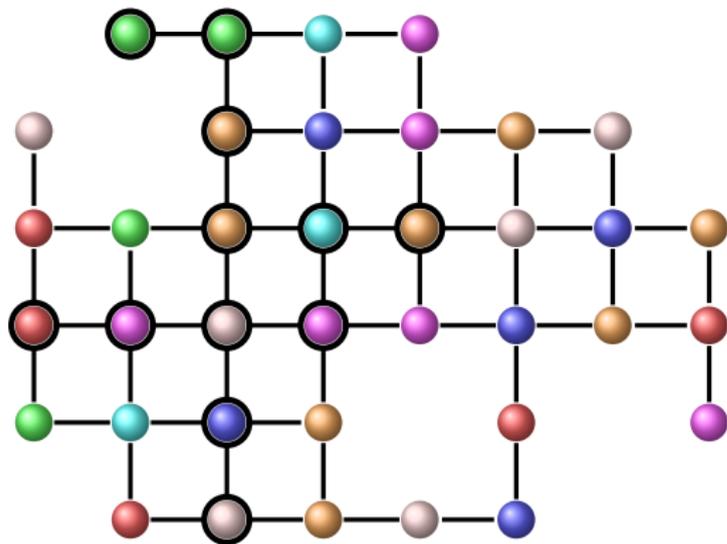


(G, λ)

MOTIF CONNEXE

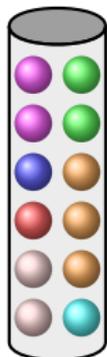


\mathcal{M}

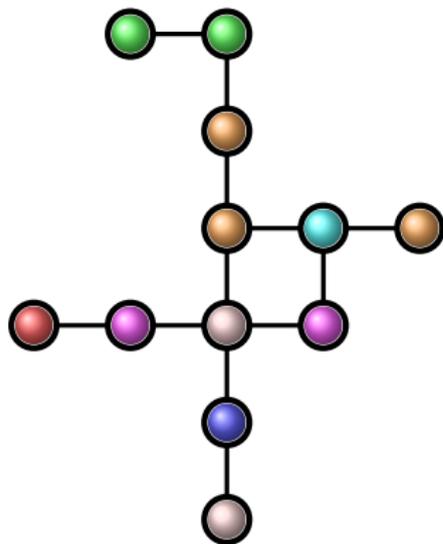


(G, λ)

MOTIF CONNEXE



\mathcal{M}



(G, λ)

MOTIF CONNEXE

- V. Lacroix, C. Fernandes, M.-F. Sagot: **Motif Search in Graphs: Application to Metabolic Networks**. IEEE/ACM Trans. Comput. Biology Bioinform. 3(4): 360-368 (2006).
- M. Fellows, G. Fertin, D. Hermelin, S. Vialette: **Sharp Tractability Borderlines for Finding Connected Motifs in Vertex-Colored Graphs**. Proceedings ICALP 2007: LNCS, 340-351.
- R. Dondi, G. Fertin, S. Vialette: **Weak pattern matching in colored graphs: Minimizing the number of connected components**. Proceedings ICTCS 2007: World Scientific, 27-38
- N. Betzler, M. Fellows, C. Komusiewicz, R. Niedermeier: **Parameterized Algorithms and Hardness Results for Some Graph Motif Problems**. Proceedings CPM 2008: LNCS, 31-43
- S. Bruckner, F. Huffner, R. Karp, R. Shamir, R. Sharan: **Topology-free querying of protein interaction networks**. Proceedings RECOMB 2009, to appear.
- R. Dondi, G. Fertin, S. Vialette: **Maximum Motif Problem in Vertex-Colored Graphs**. Submitted.

Quelques résultats

$$k = |\mathcal{M}|, n = |V| \text{ et } m = |E|$$

MOTIF CONNEXE

- MOTIF CONNEXE est résoluble en temps $O(n)$ si G est un arbre où chaque couleur apparaît au plus deux fois. Il est **NP-complet** si G est un arbre où chaque couleur apparaît au plus trois fois.
- MOTIF CONNEXE est résoluble en temps $O(2^{O(k)} n^2 \log n)$ ($O(3^k m)$ si le motif est simple).
- MOTIF CONNEXE est résoluble en temps polynomial si
 - $\text{tw}(G) = O(1)$, et
 - $|C| = O(1)$.
- MOTIF CONNEXE paramétré par $|C|$ est **W[1]-difficile**, i.e., il n'existe pas d'algorithme de complexité en temps $f(|C|) n^{O(1)}$ (pour toute fonction $f()$ ne dépendant que de $|C|$).

Motifs approchés et extensions

■ Motifs approchés

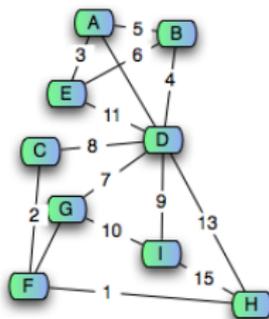
- Dispersion,
- Insertion,
- Suppression.

m i s p e l d
/ / / / / \
m i s s p e l l e d

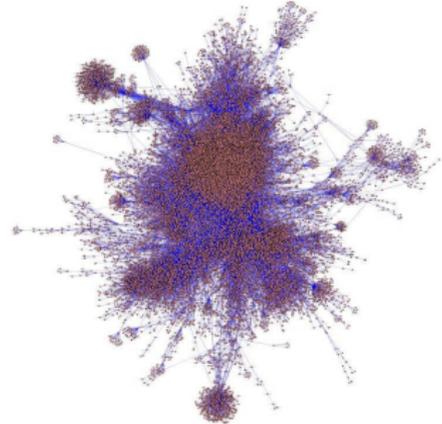
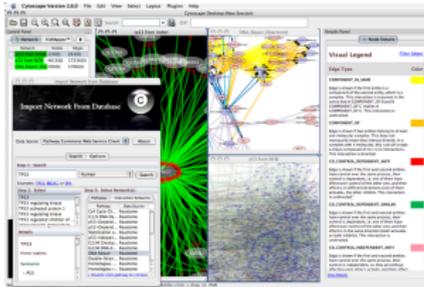


■ Extensions

- Coloration par listes
- Motifs robustes,
- Pondération.



- Recherche de motifs connexes “*approchés*”:
 - Développement d'un module pour la plateforme Cytoscape.



- Complexité paramétrée et approximation polynomial.
- Coloration par listes et sommets anonymes.