

# ANNUAIRE du **COLLÈGE DE FRANCE** 2017 - 2018

Résumé des cours et travaux

118<sup>e</sup>  
année



COLLÈGE  
DE FRANCE

— 1530 —

# INFORMATIQUE ET SCIENCES NUMÉRIQUES\*

Claire MATHIEU

Directrice de recherche au CNRS,  
professeure invitée au Collège de France

---

Mots-clés : algorithmes, données, flux, graphes

---

La série de cours et séminaires « Algorithmes » est disponible, en audio et vidéo, sur le site internet du collège de France (<https://www.college-de-france.fr/site/claire-mathieu/course-2017-2018.htm>). La leçon inaugurale *L'Algorithmique*, également disponible en audio et vidéo (<https://www.college-de-france.fr/site/claire-mathieu/inaugural-lecture-2017-2018.htm>), a été publiée sous forme imprimée (Collège de France/Fayard) et numérique (Collège de France). Texte intégral en ligne : <https://books.openedition.org/cdf/5609>.

## ENSEIGNEMENT

### COURS ET SÉMINAIRES – ALGORITHMES

Le cours a présenté un panorama de méthodes pour la conception et l'analyse d'algorithmes, en mettant l'accent sur les techniques probabilistes et l'importance de l'aléa, et avec une organisation guidée par des exemples concrets tirés de domaines divers.

### **Leçon inaugurale – L'algorithmique**

La leçon inaugurale est construite autour de quelques exemples de problèmes sociétaux : correspondances entre donneurs et malades pour organiser des greffes de reins ; placement de reines sur un échiquier ; problème du voyageur de commerce ; classement de pages web ; algorithme de Gale-Shapley et affectation de candidats à des universités ; découpage équitable de circonscriptions électorales.

---

\* Chaire créée en partenariat avec l'Inria.

## Cours 1 – Aléa et algorithmes distribués

Dans ce cours, nous avons résolu deux problèmes d’algorithmique distribuée par des algorithmes utilisant l’aléa. Tout d’abord, le problème du stable maximal a été présenté avec des applications puis résolu par l’algorithme de Luby dont l’analyse a été esquissée ; puis l’algorithme « des mouches drosophiles » a été présenté pour le même problème. Enfin, on a donné une ébauche de l’algorithme distribué pour Pagerank.

## Séminaire 1 – Algorithmes distribués

Pierre Fraigniaud (CNRS et université Paris-Diderot), le 28 novembre 2017

Ce séminaire a présenté un panorama des algorithmes distribués qui résolvent un problème *via* la coordination entre de nombreux acteurs tels des membres de réseaux sociaux, des fourmis dans une fourmilière, ou des individus dans un vol d’oiseaux migrateurs, qui doivent gérer l’incertitude temporelle ou spatiale. Concernant l’incertitude temporelle, il est impossible d’arriver à un consensus dans un cadre asynchrone, même si un seul acteur est défectueux. Il existe une caractérisation topologique des problèmes solubles dans un cadre asynchrone en présence de crashes. Concernant l’incertitude spatiale, le problème de la coloration exige du temps pour propager des informations locales, par exemple pour la coloration d’un anneau avec trois couleurs. En général, l’aléa peut aider à briser la symétrie. Plus audacieusement, on peut se poser la question d’un algorithme distribué quantique efficace pour la coloration.

## Cours 2 – Modélisation et réseaux sociaux

Ce cours a porté sur la toile (le « World Wide Web »). On y a parlé de la structure de la toile et de la distribution des degrés des nœuds, avec une queue de distribution “lourde” comprenant un nombre important de nœuds de grand degré. On a parlé de la structure des réseaux sociaux et du phénomène du petit monde. On a présenté un modèle et des résultats mathématiques pour les plafonds de verre dans les réseaux sociaux académiques.

## Séminaire 2 – Détection de communautés

Laurent Massoulié (Inria), le 5 décembre 2017

Ce séminaire a présenté des questions algorithmiques liées à la détection de communautés dans des graphes tels que les réseaux sociaux d’interaction entre personnes, les systèmes de recommandations de films à des abonnés, ou les interactions lors de réactions chimiques en biologie. L’algorithme présenté projette le graphe en petite dimension puis rassemble les nœuds proches par la méthode *k-means* pour former des groupes de similarité (*clusters*). Un modèle naturel pour simuler un graphe comprenant  $k$  communautés consiste à créer un graphe aléatoire dont les nœuds sont partitionnés en  $k$  groupes, avec probabilité d’arête plus élevée à l’intérieur qu’à l’extérieur d’un groupe. Sur un graphe créé par cette méthode, l’algorithme retrouve les  $k$  groupes si la différence entre l’intérieur et l’extérieur d’un groupe est suffisante ; en revanche, si elle est trop faible, il est impossible de reconstruire la partition.

### **Cours 3 – Théorie des jeux algorithmiques**

Deux exemples de théorie des jeux algorithmiques ont été développés dans ce cours : d'une part, le prix de l'anarchie, introduit avec l'exemple de la congestion des réseaux routiers, la notion d'équilibre de Nash, et le paradoxe de Braess. D'autre part, une introduction à la théorie des enchères a été proposée, évoquant l'algorithme de Vickrey-Clarke-Groves et l'algorithme utilisé par les moteurs de recherche pour faire payer les publicités.

### **Séminaire 3 – Calcul du prix de décisions en ligne**

Amos Fiat (université de Tel-Aviv), le 12 décembre 2017

La tarification dynamique permet d'ajuster les incitations au fil de l'eau pour refléter les changements de contexte. Un exemple est celui de la tarification des places de stationnement, où les prix peuvent augmenter lorsque le nombre de places disponibles diminue. Un autre exemple est celui de la fixation dynamique des prix pour améliorer les performances de la planification des taxis ; c'est ce que l'on appelle la tarification de la surtension, utilisée par Uber. Ce séminaire a étudié et discuté d'une variété d'applications dans lesquelles une tarification dynamique a été utilisée. Celles-ci incluent des problèmes tels que le stationnement (également appelé « correspondance métrique »), les systèmes de tâches métriques, le problème des  $k$  serveurs, les prix de compensation dynamiques du marché, la gestion des files d'attente, la modification de l'offre en réponse à une demande changeante, etc.

### **Cours 4 – Schémas d'approximation**

Ce cours a présenté des algorithmes polynomiaux et quasi-optimaux pour trois problèmes d'optimisation NP-difficiles : la découpe d'une bande de tissu en rectangles de tailles données de façon à utiliser le moins de longueur possible ; la coupe maximum dans le plan euclidien, et une modélisation de l'allocation de fréquence pour téléphonie mobile. Les techniques présentées utilisent des méthodes gloutonnes, des arrondis, et de la programmation linéaire.

### **Séminaire 4 – Combinatoire analytique**

Bruno Salvy (Inria), le 19 décembre 2017

Dans un arbre binaire aléatoire, avec quelle probabilité une feuille est-elle enfant de la racine ? Quelle est la distance moyenne d'un nœud à la racine ? Quelle en est la distribution limite ? Ce sont là des exemples typiques d'une structure combinatoire, l'arbre binaire, créé par des règles locales simples, et dont les propriétés peuvent être étudiées par l'analyse complexe. Les règles induisent des équations satisfaites par les séries génératrices, dont la méthode d'itération de Newton, associée à la transformée de Fourier rapide, permet de calculer les entrées. On peut aussi les utiliser pour l'échantillonnage de Boltzmann, et faire de l'analyse de singularités pour en déduire le comportement asymptotique. Cela se généralise à l'analyse de paramètres par l'étude de séries génératrices multivariées.

## Cours 5 – Algorithmes par marches aléatoires

Dans ce cours, nous avons étudié une forme particulière de méthodes utilisant l'aléa : les algorithmes et analyses basées sur les marches aléatoires. Une première application se trouve dans le déchiffrement de messages cryptés en remplaçant chaque lettre par une autre lettre, en modélisant la génération de tels messages par une chaîne de Markov sur les symboles, par un algorithme de recherche locale sur les solutions possibles. La méthode Métropolis a été introduite, avec application à l'autocomplétion de mots ou de phrases dans la rédaction de textos. Une deuxième application apparaît dans la génération de coloriages aléatoires de sommets d'un graphe par la méthode de Gibbs. Une troisième application se rencontre dans l'estimation de la taille d'un réseau social à partir de l'analyse des collisions lors d'une marche aléatoire dans le réseau. Dans les diverses applications esquissées, le temps de convergence est en général un paramètre important, étroitement lié à la structure du graphe exploré.

## Séminaire 5 – Échantillonnage et dénombrement approchés

Marc Jerrum (Queen Mary, university of London), le 9 janvier 2017

La complexité informatique est l'étude des ressources nécessaires pour atteindre les objectifs informatiques spécifiés. Le sujet a ses racines dans la logique, et c'est peut-être pour cela que des problèmes de décision ont généralement été au centre de l'étude, d'où notamment la théorie de la NP-complétude. Très vite, sous l'influence des applications, le spectre de la complexité informatique s'est élargi pour englober les problèmes d'optimisation.

Motivée en particulier par des applications en physique statistique et en statistique bayésienne, la complexité des calculs s'est tournée vers des problèmes d'énumération, interprétés au sens large. De tels problèmes ne peuvent que rarement être résolus exactement, et il est donc utile d'examiner la possibilité d'un comptage approximatif, ainsi que le problème de l'échantillonnage. Du point de vue algorithmique, l'échantillonnage des configurations d'un modèle physique est étroitement lié à l'estimation de sa fonction de partition. Il me semble que ce domaine est dans une situation intéressante : il y a de nombreux succès notables à célébrer, mais de nombreux problèmes non résolus, et qui intuitivement semblent étonnamment simples, restent là à nous narguer.

## Cours 6 – Algorithmes pour flux de données

Les algorithmes pour flux de données, ou algorithmes « en streaming », sont devenus populaires avec l'accroissement de la taille des données. Motivé par l'analyse du nombre de visiteurs « uniques » d'un site web, ainsi que par la détection d'attaques de type « déni de service », l'algorithme de Flajolet et Martin, permet d'estimer très rapidement, de façon fiable, et avec très peu de mémoire, le nombre de visiteurs « uniques ». Dans une autre application, on propose un algorithme simple pour vérifier qu'un fichier xml observé en mode flux de données est syntaxiquement correct. Il repose sur une propriété algébrique: un polynôme non uniformément nul, évalué en un point aléatoire, a peu de chance de donner le résultat 0. Enfin, un troisième problème est présenté, populaire sous le nom de « secretary problem » : interviewer des candidats à un poste un à un et décider si l'on recrute le candidat

interviewé avant d'avoir vu ceux qui sont encore en attente d'entretien. Un algorithme efficace exige que les candidats se présentent dans un ordre aléatoire.

### **Séminaire 6 – Algorithmes et équité**

John Kleinberg (Cornell university), le 16 janvier 2017

Une discussion récente dans la sphère publique sur la classification par des algorithmes a entraîné une tension entre des notions contradictoires sur ce que cela signifie pour une telle classification d'être *équitable* pour différents groupes. Nous examinons plusieurs des principales conditions d'équité qui sont au cœur de ces débats et discutons des recherches récentes établissant des compromis intrinsèques à ces conditions. Au cours de notre étude, nous découvrons certaines limites fondamentales du point de vue de l'équité pouvant être atteinte dans cette famille de problèmes de classification.

### **Cours 7 – Algorithmes pour graphes dynamiques**

Les graphes dynamiques évoluent au fil du temps par ajout ou suppression de nœuds ou d'arêtes. Des algorithmes sont présentés pour trois problèmes : construction en ligne d'un couplage bipartite, clôture transitive d'un graphe acyclique, et connexité de graphe.

### **Séminaire 7 – La théorie des jeux : une perspective algorithmique**

Tim Roughgarden (Stanford university), le 23 janvier 2017

Les domaines de l'informatique et de la théorie des jeux remontent tous deux à la première moitié du XX<sup>e</sup> siècle, avec les travaux d'Alan Turing, John von Neumann, et John Nash entre autres. Il existe maintenant de nombreux points de contact fructueux entre ces deux domaines : la théorie des jeux joue un rôle important dans les applications informatiques du XXI<sup>e</sup> siècle, qu'il s'agisse de réseaux sociaux ou de routage sur Internet. Le flux d'idées va également dans la direction inverse, car l'informatique offre un certain nombre d'outils qui permettent de raisonner sur les problèmes économiques de manière novatrice. Par exemple, la théorie de la complexité informatique jette un nouvel éclairage sur la « rationalité limitée » des décideurs. Les garanties d'approximation, développées à l'origine pour analyser des algorithmes heuristiques rapides, peuvent être utilement appliquées aux équilibres de Nash. Les algorithmes efficaces sont un ingrédient essentiel des conceptions modernes d'enchères à grande échelle. Dans ce séminaire ont été examinées les idées clés derrière ces connexions et leurs implications.

### **Cours 8 – Conclusion**

Ce cours, à partir de l'analyse de mutations d'ADN, donne un algorithme de reconstruction d'arbre phylogénétique. On termine par une introduction à la méthode primale-duale pour la conception d'algorithmes d'approximation, appliquée à la construction d'arbres de Steiner. Une brève conclusion résume les interactions entre problèmes, techniques et applications pour la conception et l'analyse d'algorithmes.

### Séminaire 8 – *On algorithms operating in adversarial conditions*

Allison Bishop (IEX and Columbia university), le 30 janvier 2017

Ce séminaire a discuté de la manière dont la présentation académique typique des algorithmes diffère de la réalité des algorithmes fonctionnant dans des environnements malveillants. Nous avons mis en évidence des exemples intéressants et avons commencé à travailler à une approche systématique pour combler les écarts entre la théorie et la pratique.

### COLLOQUE – *APPROXIMATION ALGORITHMS AND NETWORKS*

Le colloque « Approximation algorithms and networks » s’est tenu le 7 juin 2018 au Collège de France :

- Vincent Cohen-Addad (CNRS) : « Fast fencing » ;
- Adi Rosén (CNRS) : « Streaming algorithms for the set cover problem » ;
- Barna Saha (U. Mass Amherst) : « Space and time efficient algorithms for Lipschitz problems » ;
- Hang Zhou (École polytechnique) : « A  $(5/3+\epsilon)$ -approximation for unsplittable flow on a path: Placing small tasks into boxes » ;
- Victor Verdugo (universidad de Chile et École normale supérieure) : « Strong algorithms for the ordinal matroid secretary problem » ;
- Anna Ben-Hamou (Sorbonne université) : « Estimating graph parameters with random walks » ;
- Matthieu Latapy (CNRS) : « Stream graphs, link streams, and related algorithmic challenges » ;
- Claire Mathieu (CNRS et Collège de France) : « Hierarchical clustering: Objective functions and algorithms ».

### PUBLICATIONS

MATHIEU C., *L’Algorithmique*, Paris, Fayard/Collège de France, coll. « Leçons inaugurales du Collège de France », n° 273, 2018 ; édition électronique : Collège de France, DOI : 10.4000/books.cdf.5609.