Mathématiques discrètes : théories et usages.
Journée en hommage à Bruno Leclerc

Lundi 30 mars 2009

ENST, 46 rue Barrault, Paris 13, amphithéatre B311
et MSH 54 bd Raspail

RENSEIGNEMENTS PRATIQUES

Il est nécessaire de signaler ou de confirmer (si cela n'a déjà été fait) votre présence à tout ou partie de cette journée. Ecrire à Bernard Monjardet (Bernard.Monjardet@univ-paris1.fr)

Dans le hall d'entrée de l'ENST, la personne à l'accueil disposera de la liste des participants et vérifiera que votre nom s'y trouve (prévoir une pièce d'identité).

L'amphithéatre B311se situe près du hall d'entrée et le parcours pour y parvenir sera fléché depuis ce hall.

Pour le déjeuner, le quartier de la Butte aux Cailles dispose de nombreux restaurants (précisions en séance). Il est aussi possible de déjeuner à la cantine de l'ENST pour un tarif d'environ 11 euros. Dans ce dernier cas, le signaler dès maintenant à Olivier Hudry (hudry@enst.fr).

La journéee à l'ENST se terminera à 17h40, ce qui laissera le temps pour se retrouver à partir de **18 heures dans le hall de la Maison des Sciences de l'Homme 54 bd Raspail** pour le cocktail marquant la fin de la journée.

PROGRAMME

MATIN

9 h-9h30	Accueil, Introduction
9h30-10h10	Nathalie Caspard : LSD ou histoires de chaînes et antichaînes
10h10-10h50	Claude Barbut : Les treillis de Coxeter
11h20-12h	Patrice Bertrand : Classifications en classes empiétantes ou non et leurs dissimilarités associées
12-12h40	Bernard Fichet : Un panorama des approximations en norme du supremum pour la classification

APRES-MIDI

14h-14h15	Marc Barbut : Distributions discrètes
14h15-14h55	Alain Guénoche : Phylogénie des génomes complets et comparaison de distances entre séquences
15h15-15h55	Claude Del Vigna : Bases de données, systèmes d'implications
16h55-16h35	Florent Domenach : Hiérarchies, emboîtements, fermetures et relations bifermées
17h-17h40	Jean-Pierre Barthélemy : Approches axiomatiques et métriques en théorie formelle du consensus

18h Cocktail **EHESS, Maison des Sciences de l'Homme, 54 bd Raspail**

RÉSUMÉS des EXPOSÉS et (quelques) RÉFÉRENCES

Nathalie Caspard : LSD ou histoires de chaînes et antichaînes

LACL, Département d'informatique, Université Paris 12, 61 avenue du Général de Gaull, 94010 Creteil Cedex, ncaspard@gmail.com

Il était une fois un village pas si lointain, où l'ordre régnait, où tout était bien rangé. L'ordre était assuré - le fait est assez rare pour être mentionné – de façon bipartie par un ancien cardinal, dit "Le Clerc", et un ex-taulard dont la rencontre avec Le Clerc avait changé la vie en le libérant de ses chaînes. Cet homme, Littlewood-Offord, qui avait donc réglé ses problèmes, avait par la suite sympathisé avec un musicien, dit "Le Worth", qui recouvrait tous les soirs le village d'une douce musique, née des partitions qu'il composait lui-même.
Le produit de ces trois personnalités n'aurait été parfait s'il n'avait été également enrichi par l'arrivée du Père Nère, dit "Sperner", dont la largeur d'esprit était si étendue qu'elle foutait les Boules à bien des habitants envieux. Cet homme eut bientôt une grande famille et installa sur tous les murs extérieurs de sa maison des treillis pour y faire pousser des fleurs, symbole de liberté antichaînes. Cet homme et sa famille devinrent si aimés et populaires que les habitants rebaptisèrent le village à son nom, le "village de Sperner".
De cette histoire, vraie de toutes pièces, et de la puissante imagination de ses personnages, sont nés de jolis résultats mathématiques.
Mais ça, c'est une autre histoire.

Référence
Chapitre IV -Chaînes et antichaînes- du livre *Ensembles ordonnés finis: concepts, résultats, usages* par N. Caspard, B. Leclerc, B. Monjardet, Springer-Verlag, 2007

Claude Barbut : Les treillis de Coxeter:

CAMS 54 bd. Raspail 75270 Paris Cédex 06,barbut@ehess.fr

Cet exposé est autour d'un résultat de Bruno Leclerc sur un treillis de Coxeter partitionnable en chaînes symétriques. En 1994, Bruno Leclerc publia un exemple de treillis de Coxeter dont l'ordre n'est pas partitionnable en chaînes symétriques [1] répondant ainsi à une question de A. Bjorner [2].
Après quelques rappels sur les groupes et treillis de Coxeter (dont le permutoèdre est un exemple) et l'étude de quelques propriétés de ces treillis qui découlent de celles des groupes de Coxeter, nous rappellerons le résultat de Bruno.

Références

[1] A. Björner, "Ordering of Coxeter groups", C. Greene (ed.), *Combinatorics and Algebra. Contemporary Mathematics* 34, American Mathematical Society, Providence (R.I.), 1984, p. 175-195.

[[2] B. Leclerc, "A finite Coxeter group, the week order of which is not symmetric chain", *European Journal of Combinatorics* 15, 1994, p. 113-125.

Patrice Bertrand : Classifications en classes empiétantes ou non et leurs dissimilarités associées

Université Paris Dauphine, Ceremade, Place du Maréchal de Lattre de Tassigny, F-75775 Paris Cedex 16, & Inria Rocquencourt, AxIS., Patrice.Bertrand@ceremade.dauphine.fr

Après la bijection due à Johnson (1967) et Benzécri (1973), plusieurs auteurs ont introduit différents types de classifications en classes empiétantes (cf. E. Diday (1986), B. Fichet (1987), A. Batbedat (1988), H.-J. Bandelt et A. W. M. Dress (1989)). Ces ensembles de classifications sont en correspondance avec des ensembles de dissimilarités selon des bijections qui généralisent la bijection de Benzécri-Johnson. Dans cet exposé, nous présenterons et comparerons ces diverses bijections et plus particulièrement celles qui établissent, de façons différentes, une correspondance biunivoque entre l'ensemble de toutes les dissimilarités et des ensembles de classifications en classes empiétantes.

Références

J.-P. Benzécri et coll., *L'analyse des Données, 1 : La taxonomie*, Dunod, 1973.
A. Batbedat, Les isomorphismes HTS et HTE (après la bijection de Benzécri / Johnson), *Metro*n, 46 (1988), 47–59.
H.-J. Bandelt and A. W. M. Dress, Weak hierarchies associated with similarity measures—an additive clustering technique, *Bull. Math. Biol.*, 51 (1989), 133–166.
E. Diday, Orders and Overlapping clusters in pyramids, in : J. De Leeuw et al. (Eds.), *Multidimensional Data Analysis*, DSWO Press, Leiden, 1986, 201–234.
B. Fichet, Data Analysis : geometric and algebraic structures, in : Y. A. Prohorov, V. V. Sazonov (Eds.), *Proceedings of the First World Congress of the BERNOULLI SOCIETY,* Tachkent, 1986, V.N.U. Science Press,Vol. 2, 1987, p. 123–132.
S. C. Johnson, Hierarchical clustering schemes, *Psychometrika,* 12, n° 3, 1967, 241–254.

Bernard Fichet : Un panorama d'approximations en norme du supremum pour la classification

LIF, 163, avenue de Luminy - Case 901F-13288 Marseille Cedex 9, ,bernard.fichet@lif.univ-mrs.fr

Dans cet exposé, nous dressons un vaste panorama des approximations en norme du supremum pour nombre de structures classiques de la classification et de l'analyse des données. Seront en particulier concernées, les structures ultramétriques et hiérarchiques, leurs extensions Robinsoniennes et arborées, que ces dernières soient métriques ou de dissimilarité, ainsi que les régressions, soit convexes, soit monotones sur un ensemble partiellement

ordonné.
Tous ces résultats seront situés dans un cadre trés général, tel qu'il fut décrit dans Chepoi et al.(2000), et où les concepts de sous-dominante et de sur-dominée jouent un rôle crucial. Dans un tel cadre, toute structure sera identifiée à un sous-ensemble K d'un espace vectoriel E, p-dimensionnel, et muni de la norme du supremum, relativement a une base fixée.
Le problème d'optimisation est alors le suivant : donnés un ou n vecteurs $u_1, \cdots, u_n$ de K, trouver $\hat{y}$ de K, solution (si elle existe) de :

$$\inf_{y \in K} \max_{j=1,\cdots,n} ||y - u_j||_\infty.$$

Un vecteur u_* ($u*$) de K est appelé sous-dominante (sur-dominée) dans K de $u \in E$, s'il est le plus grand (petit) vecteur de K inférieur (supérieur) à u (au sens de l'ordre partiel défini par les coordonnées). L'existence pour tout u d'une sous-dominante (sur-dominée) revient a dire que K est un sup-(inf-) demi-treillis. On montre alors que sous cette condition, et pourvu que K soit invariant par translation le long de la diagonale principale de E, une solution du problème général est donnée par translation de la sous-dominante (sur-dominée) de l'infimum (supremum) des $u_1, \cdots, u_n$. Ces conditions sont remplies pour toutes les structures sus-évoquées, avec un calcul analytique ou algorithmique aisé d'une sous-dominante et/ou d'une sur-dominée, à l'exception des structures arborées, pour lesquelles la NP-difficulté de calcul d'une solution a été prouvée [1].
La NP-difficulté de l'approximation par une distance arborée peut être levée si l'on se restreint aux dissimilarités préservant les distances à un point donné. En effet, une transformation bien connue, dite de Farris, établit une bijection linéaire entre dissimilarités et leurs transformées, assurant qu'il y a équivalence entre la condition dite des quatre points, base des dissimilarités arborées, et ultramétricité de la transformée [3]. Sous ces conditions, il y a existence d'une sous-dominante via l'ultramétricité, et ainsi une solution au problème.

Références
1. R. Agarwala, V. Bafna, M. Farach, B. Narayanan, M. Paterson, M. Thorup, On the approximability of numerical taxonomy: Fitting distances by tree metrics. In *Proceedings of the 7th Annual ACM-SIAM Symposium on DiscreteAlgorithms*, 1996.
2. V. Chepoi, B. Fichet, l1-Approximation via Subdominants. *Journal of Mathematical Psychology*, 44: 600-616, 2000.
3. B. Leclerc, Minimum spanning trees for tree metrics: abridgment and adjustements. *Journal of Classification*, 12: 207-241, 1995.

Marc Barbut : Distributions discrètes

CAMS 54 bd. Raspail 75270 Paris Cédex 06, Marc.Barbut@ehess.fr

Il y a plusieurs façons de classer les distributions de probabilité. L'une des plus pertinentes consiste en l'existence ou non du moment d'ordre k. Une autre est l'opposition "discret-continu (dont absolument continu)". On donnera quelques indications sur chacun de ces deux points de vue.

Références
M. Barbut, *La mesure des inégalités - Ambiguïtés et paradoxes*, Droz, Genève, 2007 (seconde partie seulement).
B. Leclerc (1972), *Distributions statistiques et lois de probabilité*, coll. "Mathématiques des Sciences de l'Homme", Paris, Mouton et Gauthier-Villars.

Alain Guénoche : Phylogénie des génomes complets et comparaison de distances entre séquences

IML-CNRS, Campus de Luminy, Case 907, 13288 MARSEILLE Cedex 9
guenoche@iml.univ-mrs.fr

Bruno a très activement participé au groupe de travail sur les méthodes de reconstruction d'arbres à distances additives (1986-87) qui devait aboutir au livre *Les arbres et les représentations des proximités*. Puis nous avons étroitement collaboré sur la question de la reconstruction à partir de distances partielles (avec des valeurs manquantes). C'est dans la continuité de ces travaux que je ferai mon exposé.
La phylogénie moléculaire aborde maintenant des problèmes en vraie grandeur, puisque l'on dispose de plus de 700 séquences de génomes complets (de l'ordre de 5 mega-bases pour une bactérie, soit un livre de 500 pages). L'approche qui consiste à se fier à quelques gènes est maintenant dépassée, et les méthodes de maximum de vraisemblance basées sur les séquences alignées d'un seul gène, qui ont encore leurs adeptes, doivent être reconsidérées. Les méthodes de distance, qui commencent par mesurer des proximités entre génomes sont à nouveau d'actualité, parce qu'elles évitent le problème de l'identification des gènes communs et à cause de leur efficacité en temps de calcul. Encore faut-il que ces distances soient évolutives c'est à dire croissent avec l'éloignement au dernier ancêtre commun.
Dans cet exposé, je rapporterai les résultats d'un travail collectif qui montre que la distance basée sur les plus longs mots communs significatifs (MSM) a cette vertu et qu'elle surpasse toutes les autres distances applicables aux génomes complets.

Références

J.P. Barthélemy, A. Guénoche.- *Les arbres et les représentations des proximités*, Collection "Méthodes et Programmes", Masson, Paris, 1988, 240 p.
A. Guénoche, B. Leclerc, V. Makarenkov, On the extension of a partial metric to a tree metric, 6th International Conference on Graph Theory 2000, J.L. Fouquet, I. Rusu (Eds.) *Discrete Maths*, 276/1-3, 2003, pp. 229-248.
F. Guyon, C. Brochier, A. Guénoche, Comparison of string distances for complete bacterial genome phylogeny, IFCS'09.

Claude Del Vigna : Bases de données, systèmes d'implications

Claude Del Vigna, CAMS, claude.del-vigna@laposte.net

Un *système d'implications* sur un ensemble (fini) E est une relation binaire sur l'ensemble $P(E)$ des parties de E. On dit qu'une application f d'un ensemble O dans $P(E)$ vérifie le système d'implications R si, pour tout o dans O et pour tout (A,B) dans R, $(A \subseteq f(o)) \Rightarrow (B \subseteq f(o))$. Les exemples et les applications de ces notions sont multiples, en particulier, dans les situations où O est un ensemble d'objets, E un ensemble de propriétés et f l'application qui associe à chaque objet l'ensemble des propriétés qu'il vérifie. Les bases de données utilisent les systèmes d'implications dans au moins deux domaines : celui de la cohérence sémantique des données et celui de la fouille de données. Dans le premier cas, les systèmes d'implications sont donnés *a priori* et agissent en contrainte sur la forme des données tandis que dans le second cas, les données sont « libres » et on s'intéresse au système d'implications qu'elles vérifient. Enfin, un troisième cas plus marginal, voit apparaître les systèmes d'implications dans le rapport, qui se révèle parfois contradictoire, entre la cohérence et la confidentialité des données.
L'exposé présentera les définitions et certaines propriétés formelles des systèmes d'implication, en particulier leur lien avec les opérateurs de fermeture. Il examinera alors comment les systèmes d'implications apparaissent comme contrainte dans les bases de données relationnelles – sous la forme d'ensemble de dépendances fonctionnelles – et comment ils constituent un élément du processus de décomposition des tables relationnelles. Il exposera ensuite l'occurrence des systèmes d'implications dans la fouille de données et l'importante notion de base de règles. Enfin, il présentera une situation où cohérence et

confidentialité des données « jouent l'une contre l'autre » en faisant apparaître un système d'implications.

Références
[1] S. Abiteboul, R. HullL, V. Vianu,. *Foundations of databases*. Addison-Wesley, 1995.
[2] N. Caspard, B. Leclerc, B. Monjardet. *Ensembles ordonnés finis : concepts, résultats et usages*. Springer, 2007.
[3] C. Delobel, M. Madiba.*Bases de données et systèmes relationnels*. Dunod-Informatique, 1982.

Florent Domenach : Hiérarchies, emboîtements, fermetures et relations bifermées

Computer Science Department, Université de Nicosie, 46 Makedonitissas Av., 1700 Nicosia, Cyprus ,domenach.f@unic.ac.cy

Les systèmes de fermeture (ou familles de Moore) sont liées par cryptomorphismes a de nombreuses structures [1], parmi lesquels (pour les plus célèbres) on peut citer les opérateurs de fermeture et les systèmes implicatifs complets. Nous allons présenter ici une structure particulière, les relations d'emboîtement, telles qu'introduites dans [2]. Nous allons montrer que ces relations d'emboîtements permettent tout à la fois de caractériser différents systèmes de fermeture et de généraliser des résultats d'ajustement de classifications.

Références
[1] Caspard N., Monjardet B., The lattices of Moore families and closure operators on a finite set: a survey, *Discrete Applied Math*.(2003) 127, 241-269.
[2] Domenach F., Leclerc B., Closure systems, implicational systems, Overhanging relations and the case of Hierarchical Classification, *Mathematical Social Science* (2004) 47 (3), 349-366.

Jean-Pierre Barthélemy : Approches axiomatiques et métriques en théorie formelle du consensus

Cet exposé va porter sur des travaux de Bruno et quelques autres (Bernard Monjardet, Bill Day, Buck Mc Morris….) sur les problèmes de consensus en classification. Il devrait également s'étendre sur diverses approches métriques auxquelles Bruno a énormément participé.

Références
J.-P. Barthélemy and B. Leclerc (1995) The median procedure for partitions. In P. Hansen (ed) *Partitioning data sets*, American Mathematical Society,Providence, 3-34.
O. Hudry, B. Leclerc, B. Monjardet et J.P. Barthélemy Médianes métriques et latticielles, chapitre 6 du volume 3 du traité *Méthodes pour l'aide à la décision,* (eds Denis Bouyssou, Didier Dubois, Marc Pirlot et Henri Prade), pp 274-279 Hermès, Paris, 2006.
W.H.E. Day, F.R. McMorris, *Axiomatic Consensus Theory in Group Choice and Biomathematics*, SIAM, Philadelphia, 2003.
B. Leclerc, B. Monjardet (1995), Latticial theory of consensus, in *Social choice, Welfare and Ethics*, V. Barnett, H. Moulin, M. Salles & N. Schofield (eds), Cambridge University Press, 145-159.