n^{n-2}

PAR

PIERRE DUCHET

0. INTRODUCTION.

La formule de Cayley [3] qui évalue le nombre d'arbres étiquetés à n sommets reçut en 1918 une jolie démonstration bijective due à Prüfer [11], qui fournit un codage canonique de ces arbres dans l'ensemble produit $[n]^{n-2}$, où $[n] := \{1, \ldots, n\}$. Une généralisation de ces résultats est ici proposée: Cayley et Prüfer énuméraient les arbres qui assurent la connexion des éléments d'un unique ensemble; on obtient ici le codage et l' énumération des arbres sur $[n]$ qui assurent la connexion des éléments de chaque membre d'une famille quelconque $X_1, X_2, \ldots, X_m$ de parties de $[n]$.

Dans tout l'article, X désignera l'ensemble générique à n éléments. Un *hypergraphe sur* X est un couple $\mathcal{H} = (X, \mathcal{F})$ où $\mathcal{F}$ est une famille de parties de X (cf. [2] et [9]). Les ensembles de $\mathcal{F}$ sont les *arêtes* de $\mathcal{H}$; les éléments de X sont les *sommets* de $\mathcal{H}$ et n est l'*ordre* de $\mathcal{H}$. D'autres notions indispensables seront introduite au moment de leur utilisation.

Evidemment, le problème considéré n'a d'intérêt que pour les hypergraphes qui admettent au moins un arbre solution. Connus sous le nom d'*hypergraphes arborés*, ces hypergraphes ont été caractérisées de diverses manières [4,8,12]. Voir [2,5,6] pour une approche synthétique de ces structures.

(0.1) Definition. On note $\Pi(X)$ l'ensemble des arbres dont les sommets sont les éléments de X. Un arbre $T \in \Pi(X)$ est un *arbre de base* pour un hypergraphe $\mathcal{H} = (X, (X_i)_{i \in I})$ si chaque X_i induit un sous arbre de T. On note $\Pi(\mathcal{H})$ l'ensemble des arbres de base de $\mathcal{H}$. L'hypergraphe $\mathcal{H}$ est *arborée* dès que $\Pi(\mathcal{H})$ est non vide.

1. CAS D'UNE PARTITION: SOUS-ARBRES PRESCRITS.

Soit $\mathcal{M} = \{X_1, \ldots, X_m\}$ une partition de X. Par commodité on identifiera $\mathcal{M}$ à l'hypergraphe $(X, \mathcal{M})$. Tout arbre de $\Pi(\mathcal{M})$ peut être construit en choisissant d'abord pour chaque i, $1 \leq i \leq m$, un arbre $T_i \in \Pi(X_i)$, puis en complétant la forêt obtenue. Notons $\Pi(X; T_1, T_2, \ldots, T_m)$ l'ensemble des arbres de $\Pi(X)$ qui contiennent chaque T_i.

(1.1) Proposition. (*Moon* [10]). $|\Pi(X; T_1, T_2, \ldots, T_m)| = |X_1| \cdots |X_m| \cdot |X|^{m-2}$. □

Ainsi la cardinalité de $\Pi(X; T_1, T_2, \ldots, T_m)$ ne dépend que de la partition $\mathcal{M}$. En notant $\Pi_{\mathcal{M}}$ l'ensemble $X_1 \times \cdots \times X_m \times X^{m-2}$, on a:

(1.2) Corollaire. *Pour toute partition $\mathcal{M} = (X_i)_{i \in I}$ de X, on a:*

$$\Pi(\mathcal{M}) \simeq \Pi_{\mathcal{M}} \times \prod_{i \in I} \Pi(X_i). \qquad \square$$

Nous allons en fait redémontrer la formule de Moon en exhibant une bijection canonique entre les ensembles $\Pi(X; T_1, T_2, \ldots, T_m)$ et $X_1 \times \cdots \times X_m \times X^{m-2}$, dont la construction s'inspire du codage de Prüfer.

Soit $T \in \Pi(X; T_1, T_2, \ldots, T_m)$; en contractant chaque T_i en un sommet unique, l'arbre T devient un arbre à m sommets que l'on peut identifier à un arbre T_* sur $\{1, \ldots, m\}$. Définissons une séquence d'entiers $r(1), \ldots, r(m)$ et une séquence $p(1), \ldots, p(m)$ d'éléments de X de la maniére suivante:

$$\text{(1.3)} \qquad \begin{cases} r(1) \text{ est le plus petit sommet pedant de l'arbre } T_*; \\ r(i+1) \text{ est le plus petit sommet pendant de l'arbre } T_* \backslash \{r(l), \ldots, r(i)\}; \end{cases}$$

$$\text{(1.4)} \qquad \begin{cases} p(i) \text{ est l'unique sommet du sous-arbre } T \backslash (X_{r(1)} \cup \cdots \cup X_{r(i)}) \text{ qui est} \\ \quad \text{adjacent (dans } T\text{) à l'ensemble } X_{r(i)}, \text{ ceci pour chaque } i \in \{1, \ldots, m\}. \end{cases}$$

Remarquons que $p(m-1) \in X_m$ et que r est une permutation de $[m]$ qui laisse m invariant. Définissons une autre séquence d'éléments de X, soit $p_1, \ldots, p_m$ par les conditions suivantes:

$$\text{(1.4)} \qquad \begin{cases} \text{Pour } 1 \leq i \leq m-1, \{p_{r(i)}, p(i)\} \text{ est la seule arête de } T \\ \qquad \text{qui sorte del'ensemble } X_{r(i)}; \\ p_m := p(m-1). \end{cases}$$

(1.5) Theorème et définition. *Soit $\mathcal{M} = (X_1, \ldots, X_m)$ une partition de X, et pour $1 \leq i \leq m$ soit T_i un arbre sur X_i. L'applicaton p_* qui à chaque arbre T de $\Pi(X; T_1, \ldots, T_m)$ associe son code de Prüfer généralisé, défini par:*

$$p_*(T) = (p_1, p_2, \ldots, p_m, p(1), p(2), \ldots, p(m-2))$$

est une bijection de $\Pi(\mathcal{M})$ sur l'ensemble $X_1 \times \cdots \times X_m \times X^{m-2}$.

Indication sur la démonstration: il suffit de vérifier que la suite $r(i)$ est caractérisée par la propriété:

$$\text{(1.6)} \qquad \begin{cases} r(i+1) \text{ est le plus petit entier qui ne figure pas dans la suite} \\ r(1), \ldots, r(i), p(i+1), \ldots, p(m-1), \end{cases}$$

ce qui permet de construire p_*^{-1}. On remarquera que la séquence des $p(i)$, avec $1 \leq i \leq m-2$, est le code de Prüfer usuel correspondant à l'arbre T_*. $\square$

2. FAMILLES ARBOREES, CONVEXES ET LIENS.

Deux idées essentielles gouvernent l'étude de la structure des familles arborées. La première est la remarque que toute intersection de parties connexes d'un arbre est encore connexe; ainsi les parties qui induisent des sous-arbres constituent les fermés d'une "fermeture de Moore" ou d'une "convexité" au sens abstrait (cf. [7] pour ce point de vue). La seconde idée est que parmi ces "convexes", certains jouent un rôle essentiel dans la connexion de la famille arbrée considérée. Les définitions suivantes sont relatives à un hypergraphe sur X, soit $\mathcal{H} = (X, (X_i)_{i \in I})$.

(2.1) Definition. Les parties de X de la forme $\bigcap_{i \in J} X_i$, où $J \subseteq I$, sont appelés parties *$\mathcal{H}$-convexes.* En particulier l'ensemble X est convexe. L'*enveloppe convexe* d'une partie A de X est l'intersection des convexes qui contiennent A; elle est notée $[A]_{\mathcal{H}}$.

(2.2) Definition. La *localisation* de $\mathcal{H}$ sur une partie Y de X, notée $\mathcal{H}]Y[$ est la famille des traces sur Y des ensembles de $\mathcal{H}$ qui ne contiennent pas Y. On appelle *lien de* $\mathcal{H}$ toute partie Y de X qui est $\mathcal{H}$-convexe et telle que la localisation $\mathcal{H}]Y[$ soit non-connexe; il est commode de considérer qu'un hypergraphe vide d'arêtes est non-connexe; ainsi chaque singleton $\{x\} \subseteq X$ est-il un lien, dit *trivial.* Les composantes connexes de $\mathcal{H}]Y[$ forment ce qu'on appelle la *partition (de liasion)* de Y relative à $\mathcal{H}$, notée $\mathrm{Part}_{\mathcal{H}}(Y)$ ou simplement $\mathrm{Part}(Y)$.

EXEMPLES. Les arêtes d'un graphe G sont des liens de G (la partition de liaison associée à chaque arête est la bipartition triviale en singletons). Si G est connexe, il n'y a pas d'autres lien non trivial. Si G est non-connexe, son ensemble de sommet est aussi un lien. Si $\mathcal{F}$ est une partition de X, l'ensemble X lui-meĥe est un lien de l'hypergraphe $\mathcal{H} = (X, \mathcal{F})$ et $\mathcal{F}$ est la partition de liaison correspondante; les autres liens non triviaux de $\mathcal{H}$ sont les ensembles de la partition $\mathcal{F}$ (les partitions de liaisons correspondantes étant les partitions triviales en singletons).

(2.3) Proposition. *Tout lien non trivial L d'un hypergraphe $\mathcal{H}$ est l'enveloppe $\mathcal{H}$-convexe d'un ensemble de cardinal 2.* $\square$

(2.4) Proposition. *Un hypergraphe arboré d'ordre n a au maximum $n-1$ liens non triviaux.* $\square$

3. DETERMINATION DES ARBRES DE BASE.

La structure de liasion d'un hypergraphe $\mathcal{H}$ sur X est l'ensemble de ses liens, considéré comme hypergraphe sur X et muni de la relation d'ordre $\subseteq$. Lorsque la structure de liasion est considérée comme hypergraphe, les liens triviaux ne jouent aucun rôle. Les propositions précédentes indiquent que la structure de liaison est plus simple (remarquablement plus simple, en général) que l'hypergraphe. Néanmoins, pour le problème qui nous interesse ici, l'étude de la structure de liaison suffit:

(3.1) Lemma fondamental. *Soit $\mathcal{L}$ la structure de liaison d'un hypergraphe arboré $\mathcal{H}$. On a:* $\Pi(\mathcal{L}) = \Pi(\mathcal{H})$. $\square$

(3.2) Théorème. *Soit $\mathcal{H}$ un hypergraphe arboré sur X et soit $\mathcal{L}$ l'ensemble de ses liens. Il existe une bijection canonique de $\Pi(\mathcal{H})$ sur l'ensemble produit*

$$\prod_{L \in \mathcal{L}} \Pi_{Part(L)}.$$

(3.3) Corollaire. *Soit $(L_i)_{1 \leq i \leq m}$ la liste des liens de l'hypergraphe arboré $\mathcal{H}$. Chaque lien L_i a l_i sommets, $c(i)$ composantes connexes de cardinalités respectives $c_{i,j}$, $1 \leq j \leq c(i)$. Alors, les nombres d'arbres de base de $\mathcal{H}$ est donné par la formule:*

$$|\Pi(\mathcal{H})| = \prod_{1 \leq i \leq m} l_i^{c(i)-2} \prod_{1 \leq j \leq c(i)} c_{i,j}.$$

Quelques indications sur la démonstration: en fait tout arbre de base d'un hypergraphe arboré $\mathcal{H}$ peut être construit par extensions successives, en choisissant de proche en proche un arbre sur chaque lien, obtenu par extension des arbres déjà contruits sur les liens plus petits. Il s'avère que les choix d'extensions possibles sont mutuellement indépendants (cela résulte notamment des résultats énoncés en annexe); comme chaque extension est un cas où s'applique le théorème (1.5), on aboutit au théorème (3.2) et à son corollaire. Pour plus de détails voir [6]. $\square$

4. COMMENTAIRES ET APPLICATIONS.

(4.1) Il n'est pas très difficile de bâtir un algorithme qui, étant donné un hypergraphe arboré $\mathcal{H}$ à n sommets et m arêtes, détermine en temps $O(mn^3)$ la structure de liaison et par là, l'ensemble des arbres de base. Le nombre d'arbre de base d'un hypergraphe arboré $\mathcal{H}$ est en général exponentiellement grand par rapport à m et n, il est donc clair que le comptage et la description complète de $\Pi(\mathcal{H})$ ne peut se faire de manière rapide qu'en raison de la structure produit (de longueur polynomiale en n) de l'ensemble $\Pi(\mathcal{H})$, énoncée par le théorème (3.2). A tout arbre de base peut être canoniquement associé un code "de Prüfer généralisé" dans l'ensemble produit

$$\prod_{L \in \mathcal{L}} \Pi_{Part(L)}.$$

Ce code s'obtient par concaténation des codages de Prüfer généralisés associés, selon le théorème (1.5), aux sous-arbres induits par les liens modulo les partitions de liaison correspondantes.

(4.2) Le concept de structure de liaison rend possible la solution d'un vieux problème (dû à Renz, cf. [6]) à savoir la caractérisation des hypergraphes de chemins d'arbre, c'est à dire des hypergraphes arborés admettant un arbre de base T qui induise un chemin sur chaque arête de $\mathcal{H}$. Une telle caractérisation, de type "configurations exclues", a récemment été obtenue par Arami et l'auteur [1], aussi bien dans le cas orienté que non orienté. La démonstration de ces caractérisations permet la reconnaissance de ces hypergraphes par un algorithme rapide, basé sur l'algorithme (4.1).

(4.3) Plus générallement, les méthodes développées ici devraient permettre une organisation mieux structurée des bases de données relationnelles existantes, puisque l'efficacité d'une telle organisation repose, en l'état actuel de la question, sur les hypergraphes arborés (ou leur duaux) et sur la construction d'arbres de base d'un type donné pour ces hypergraphes.

(4.4) Il est possible d'interpréter les nombres d'arbres de base d'un hypergraphe arboré comme les déterminants d'un certain type de matrices qui, pour des raisons non encore élucidées, interviennent également en Géométrie Algébrique ("M-matrices de graphes algébriques"). Il semble intéressant dans ce contexte de développer des généralisations matricielles du théorème (3.2), par analogie avec les résultats classiques sur l'énumération des arbres couvrants d'un graphe donné. Cela sera l'objet d'un travail ultérieur.

ANNEXE: LIENS ET PARTITTONS SPECTRALES.

Nous allons donner une autre description de la structure de liaison d'un hypergraphe arboré, utile du point de vue algorithmique. Dans toute cette section $\mathcal{H}$ désigne un hypergraphe sur X. Si σ est une partition quelconque de X, on note $\mathcal{H}/\sigma$ l'hypergraphe quotient obtenu par contraction de chaque ensemble constitutif de σ. Pour toute partition partielle $\tau \subseteq \sigma$, on note $\hat{\tau}$ l'intersection de tous les ensembles $\mathcal{H}$-convexes qui rencontre chaque ensemble de τ.

(A.1) Definition. Le *spectre (d'ordre* 1*)* d'un hypergraphe $\mathcal{H}$ sur X, noté $\mathrm{sp}(\mathcal{H})$ est l'hypergraphe sur X formé des liens non triviaux minimaux (pour l'inclusion). La *séquence spectrale* $S_1, \ldots, S_s$ de $\mathcal{H}$ est définie récursivement par:

$$\text{(A.2)} \qquad \begin{cases} S_1 := \mathrm{sp}(\mathcal{H}) \\ S_{k+1} := \mathrm{sp}(\mathcal{H}/\sigma_k) \text{ oú } \sigma_k \text{ est la partition de } X \text{ formée} \\ \qquad\qquad \text{des composantes connexes de } S_k. \end{cases}$$

Les σ_k sont appelées les *partitions spectrales* de $\mathcal{H}$. L'entier s (la "hauteur" de $\mathcal{H}$) est le premier indice tel que S_{s+1} ne contienne aucune arête. La démonstration purement technique du lemme suivant est laissé au soin du lecteur:

(A.3) Lemme. *Si $\mathcal{H}$ est arboré, chaque spectre S_k (spectre de hauteur k) est une hyperforêt, c'est à dire un hypergraphe sans aucun cycle. L'application $E \mapsto \hat{E}$ est une bijection entre l'ensemble des arêtes des spectres successifs de $\mathcal{H}$ et l'ensemble des* *liens non triviaux de $\mathcal{H}$. Une arête de E de S_k est l'image du lien $\hat{E}$ par la surjection canonique $X \to X/\sigma_k$; la partition σ_k correspond par cette surjection à la partition de liaison de $\hat{E}$.*

Une description des arbres de base de $\mathcal{H}$ et une démonstration du théorème (3.2) sont obtenus à partir du résultat suivant où σ_0 désigne la partition triviale de X en singletons:

(A.4) Théorème. *Si $\mathcal{H}$ est un hypergraphe arboré, les arbres de base pour $\mathcal{H}$ sont exactement les arbres T sur X tels que chaque quotient T/σ_{k-1} soit, pour $1 \leq k \leq s$, un arbre de base de S_k.*

References

[1] Z. Arami, P. Duchet, *Systems of paths in a tree*, en preparation.
[2] C. Berge, *Hypergraphes*, Gauthiers-Vilars, Paris, 1987.
[3] A. Cayley, *A theorem on trees*, Quart. J. Math. **23** (1889), 376–378.
[4] P. Duchet, *Propriété de Helly et problèmes de représentation*, Proc. Int. Coll. CNRS **260** (1978), CNRS, Paris, 117–118.
[5] P. Duchet, *Classical Perfect Graphs, an introduction with emphasis on triangulated and interval graphs*, Ann. Disc. Math. **21** (1984), 67–96.
[6] P. Duchet, *Tree hypergraphs and their basic trees*, Preprint, Univ. Paris 6, 1985.
[7] P. Duchet, *La convexité dans les structures combinatoires*, Actes du Sémin. Lotharingien, 12e Session, Kleebach, 1985, pp. 67–107.
[8] C. Flament, *Hypergraphes arborés*, Discrete Math. **21** (1978), 223–226.
[9] L. Lovász, *Combinatorial problems and exercises*, North Holland Pub. Co., Amsterdam–New York–Oxford, 1979.
[10] J. W. Moon, *Enumerating labelled trees*, Graph Theory and Theoretical Physics (F. Harary, ed.), Acad. Press, London–New York, 1967, pp. 261–271.
[11] A. Prüfer, Archiv Math. Phys. **27** (1918), 142–144.
[12] P. J. Slater, *A characterization of soft hypergraphs*, Canad. Math. Bull. **21** (1978), 335–337.