BULLETIN DE LA S. M. F.

CHRISTOPHE REUTENAUER

Ensembles libres de chemins dans un graphe

Bulletin de la S. M. F., tome 114 (1986), p. 135-152

<http://www.numdam.org/item?id=BSMF_1986__114__135_0>

Bull. Soc. math. France,
114, 1986, p. 135-152.

ENSEMBLES LIBRES DE CHEMINS DANS UN GRAPHE

PAR

CHRISTOPHE REUTENAUER (*)

RÉSUMÉ. — On établit les bases d'une extension de la théorie des codes à longueur variable aux chemins dans les graphes orientés : algorithme de codicité, libérabilité, maximalité et complétude, factorisation. On caractérise les graphes fortement connexes par l'irréducibilité de leur déterminant.

ABSTRACT. — We lay down the basis of an extension of the theory of variable length codes to paths in directed graphs: algorithm of codicity, liberability, maximality and completeness, factorization. We characterize strongly connected graphs by the irreducibility of their determinant.

1. Introduction

Le but de cet article est de démontrer quelques éléments d'une théorie des ensembles libres de chemins dans un graphe, considérée comme une extension de la théorie des codes à longueur variable,

Nous appelons *code* un ensemble de chemin dans un graphe orienté qui satisfait la condition de décomposition unique : tout chemin du graphe se décompose en au plus une manière comme produit de chemins du code. Le cas des codes de mots (codes à « longueur variable ») se retrouve en considérant des graphes à un sommet.

Nous commençons par passer en revue les résultats de base en théorie des codes (parag. 3) : l'algorithme de Sardinas et Patterson pour tester si

(*) Texte reçu le 28 février 1985.

C. REUTENAUER, Institut de Programmation, 4, place Jussieu, 75231 Paris Cedex 05.

BULLETIN DE LA SOCIÉTÉ MATHÉMATIQUE DE FRANCE — 0037-9484/1986/02 135 18/$3.80

un ensemble de chemins est un code, la condition de codicité en termes de séries caractéristiques, la condition de libérabilité d'une sous-catégorie ([1]) et son corollaire sur la liberté d'une intersection de sous-catégories libres, enfin un exemple spécifique aux graphes qui servira plus loin.

Au paragraphe 4, nous étudions les liens entre codes complets, codes maximaux et codes complets minimaux. Il n'y a plus identité entre ces trois notions, comme c'est le cas pour les codes de mots d'après un théorème de Schützenberger. Mais nous caractérisons les graphes où l'on a équivalence entre ces trois notions (th. 1) : ce sont les graphes possédant la propriété

(0) autour de chaque sommet, il y a au moins une boucle

et nous caractérisons aussi les codes complets minimaux (th. 2).

Au paragraphe 5, nous démontrons un résultat de factorisation du déterminant associé à un code complet (th. 4) : celui-ci est divisible par le déterminant du graphe. Nous le prouvons sous l'hypothèse (0), et conjecturons le résultat en général. Cette conjecture est un cas particulier d'une autre, plus générale (parag. 6) : la matrice du code est un multiple de la matrice du graphe. C'est un problématique analogue à celle soulevée par la conjecture de factorisation des codes de mots (Schützenberger, Césari, Perrin). Nous démontrons, comme résultat préliminaire au théorème 4, un résultat intéressant en soi : le déterminant du graphe est un polynôme irréductible si et seulement si le graphe est fortement connexe (sauf cas triviaux) (*cf.* th. 3).

2. Motivations ([2])

Un *code* (à longueur variable) est une partie libre d'un monoïde libre (i. e. la base d'un sous-monoïde libre). Un exemple classique est le code

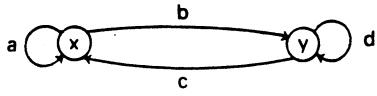
---

([1]) L'objet où vivent les codes de chemins est une catégorie libre, par opposition aux codes de mots, où c'est le monoïde libre. Étudier les codes de chemins revient à étudier les sous-catégories libres des catégories libres.

([2]) Celles-ci sont issues d'une suggestion de M. P. Schützenberger, lors de discussions entre lui et l'auteur.

$$X=\{xx,\ xy,\ xxy,\ xyy,\ yy\},$$

contenu dans le monoïde libre $(x, y)^*$ à deux générateurs; on sait que X est un code *maximal* (i. e. tout sur-ensemble n'est plus un code). Chaque mot w en les lettres x, y peut être considéré comme un chemin dans le graphe G suivant



chemin d'origine la première lettre de w et déterminé par la suite des lettres dans w; par exemple, le mot $w=xxyxyyy$ détermine le chemin

$$\begin{matrix} x & x & y & x & y & y & y \\ a & b & c & b & d & d & \end{matrix}$$

selon la règle : $xx \to a$, $xy \to b$, $yx \to c$, $yy \to d$.

Le code X ci-dessus détermine donc les chemins

$$C'=\{a,\ b,\ ab,\ bd,\ d\}.$$

Mais un mot dans X^* (le sous-monoïde engendré par X), i. e. de la forme $u_1 u_2 \ldots u_n$, $u_i \in X$, détermine un chemin qui n'est pas $c_1 c_2 \ldots c_n$, où c_i est déterminé par u_i : il manque les transitions entre les x_i successifs; c'est pourquoi on considère, au lieu de C, l'ensemble de chemins $C=C'.\{a, b, c, d\}$, où le produit (partiel) est celui des chemins, i. e.

$$C=\{aa,\ ab,\ bc,\ bd,\ abc,\ abd,\ bdc,\ bdd,\ dc,\ dd\}.$$

Il est naturel de lui associer la matrice, qui repère pour chaque chemin ses deux extrémités :

$$M=\begin{pmatrix} a^2+bc+abc+bdc & ab+bd+abd+bd^2 \\ dc & d^2 \end{pmatrix}$$

Le calcul montre que

$$\det(I\text{-}M)=1-a^2-bc-d^2-abc-bcd-abcd+a^2d^2,$$

qui se factorise sous la forme :

$$(1-a-d+ad-bc)(1+a+d+ad).$$

Mais le premier facteur est $\det(I-N)$, où

$$N=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

est la « matrice du graphe ».

C'était ce genre de résultat qu'il s'agissait de prouver.

En fait, la codicité de X implique que tout chemin dans G qui se décompose comme produit de chemins dans C admet une unique décomposition de cette forme.

Il est alors tentant d'appeler X un *code* de chemins dans G; et de considérer des graphes quelconques en place du graphe particulier G. De plus, X étant complet (puisque maximal : théorème de Schützenberger), i. e. tout mot sur $\{x, y\}^*$ est facteur d'un mot dans X^*, l'ensemble C de chemins jouit d'une propriété analogue : tout chemin dans G se prolonge (à gauche et à droite) en un chemin de la forme $c_1 c_2 \ldots c_n$, $c_i \in C$.

On dira donc d'un tel code de chemins qu'il est *complet*. Le résultat ci-dessus, que $\det(I-M)$ est divisible par $\det(I-N)$ pour un code complet C, s'interprète alors comme une généralisation d'un résultat classique pour les codes de mots (qui remonte à Shannon) : si X est un code maximal fini, alors $1-\mathbf{X}$ est divisible par $1-\mathbf{A}$ ($\mathbf{X}$ désigne le polynôme somme des monômes qui sont image commutative des mots dans X).

3. Généralités

Nous considérons un graphe (orienté) $G=(S, A)$ où S est l'ensemble (fini) des sommets et A l'ensemble (fini) des arcs. Nous notons $\mathscr{C}$ l'ensemble des chemins du graphe, y compris les chemins vides $1_s : s \to s$, pour tout sommet s. En fait, $\mathscr{C}$ est une catégorie, dont les objets sont les sommets du graphe, et les morphismes les chemins : $\mathscr{C}$ est la *catégorie libre engendrée* par G.

Chaque chemin c non vide est représenté par un mot non vide sur l'alphabet A, que nous notons encore c (mais chaque mot ne représente pas forcément un chemin, sauf si $|S|=1$).

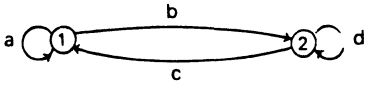
Nous appellerons *code* tout ensemble C de chemins tel que : pour tous $c_1, \ldots, c_n, d_1, \ldots, d_p$ dans C, la relation $c_1 c_2 \ldots c_n = d_1 d_2 \ldots, d_p$ (où il est entendu que ces produits sont définis, i. e. sommet terminal (c_{i-1}) = sommet initial (c_i) etc.) implique $n=p$ et $c_i=d_i$, $\forall i$. Autrement dit,

si un chemin dans G se décompose en un produit de chemins dans C, alors cette décomposition est unique. Nous notons C^* l'ensemble de ces chemins, y compris tous les chemins vides 1_s, $s \in S$. Une définition alternative est de dire que C est un code si et seulement si C^* est la catégorie libre engendré par le graphe $G(C)=(S, C)$ ([3]).

Exemple

$$S=\{1, 2\}, \qquad A=\{a:1 \to 1,\ b:1 \to 2,\ c:2 \to 1,\ d:2 \to 2\}.$$

On remarquera que conformément à une remarque précédente, le mot ad (par exemple) ne représente pas un chemin.



C est l'ensemble des 10 chemins $\{ aa, bc, abc, bdc, ab, bd, abd, bdd, dc, dd \}$. Pourquoi C est-il un code? Cela découle d'une généralisation de l'algorithme de Sardinas et Patterson (*voir* par exemple [4], prop. 5.2.7) : si E, F sont des ensembles de chemins, on note

$$E^{-1}F=\{d \mid \exists c \in E,\ cd \in F\}.$$

Nous notons $I=\{1_s \mid s \in S\}$ l'ensemble des chemins vides. Soient alors

$$X_0=C^{-1}C \setminus I,$$

$$X_{i+1}=X_i^{-1}C \cup C^{-1}X_i \quad (i \geqslant 0).$$

Alors C *est un code si et seulement si aucun des* $X_n (n \geqslant 0)$ *ne contient de chemin vide* : ceci se démontre de manière tout à fait analogue au cas des codes de mots, et de ce fait nous omettons la preuve. Il est clair, de plus, que si C est fini, alors les longueurs des chemins dans les X_n décroissent, donc la suite (X_n) est fini ou périodique, et l'on obtient ainsi un algorithme effectif pour tester si C est un code ou non. Dans l'exemple ci-dessus, on a $X_0=\{c, d\}$, $X_1=\{c, d\}=X_2$, etc. On est donc bien en présence d'un code.

([3]) On retrouve la notion usuelle de codes [1] (à longueur variable; nous dirons : codes de mots) en se restreignant aux graphes à un sommet; en effet dans ce cas, l'ensemble des chemins s'identifie au monoïde libre A^* engendré par A (le produit de deux chemins est toujours défini).

Nous considérons maintenant l'algèbre libre $\mathbb{Z}\langle A\rangle$ des polynômes non commutatifs engendrés par A sur $\mathbb{Z}$. De même, l'algèbre $\mathbb{Z}\langle\langle A\rangle\rangle$ des séries formelles non commutatives engendrée par A sur $\mathbb{Z}$. Chaque élément de $\mathbb{Z}\langle A\rangle$ (resp. $\mathbb{Z}\langle\langle A\rangle\rangle$) est une combinaison linéaire (resp. combinaison linéaire infinie) de mots sur l'alphabet A. Nous notons $\mathbb{Z}\langle A\rangle^{S\times S}$ (resp. $\mathbb{Z}\langle\langle A\rangle\rangle^{S\times S}$) l'algèbre des matrices carré de taille $S\times S$ à coefficients dans $\mathbb{Z}\langle A\rangle$ (resp. $\mathbb{Z}\langle\langle A\rangle\rangle$). A chaque chemin non vide $c : s\to t$, nous associons la matrice $\mathbf{c}=cE_{s,t}$ (i.e. la matrice ayant comme coefficient de coordonnées (s, t) le mot c, les autres coefficients étant nuls); au chemin vide 1_s, nous associons $\mathbf{1}_s=E_{s,s}$. Nous appelons *algèbre du graphe* G, notée $\mathscr{A}$, la sous-algèbre de $\mathbb{Z}\langle A\rangle^{S\times S}$ formée des éléments de $\mathbb{Z}\langle A\rangle^{S\times S}$ qui sont combinaison linéaire des $\mathbf{c}$, c chemin dans G ($\mathscr{A}$ n'est autre qu'une algèbre de carquois, ou « quiver-algebra » au sens de P. Gabriel, *cf.* par exemple [3]).

De même, nous notons $\mathscr{S}$ la sous-algèbre de $Z\langle\langle A\rangle\rangle^{S\times S}$ des combinaisons linéaires infinies de c : nous l'appellerons l'*algèbre large du graphe*. A chaque ensemble de chemin L, nous associons sa *série caractéristique*, notée $\mathbf{L}$ et définie par $L=\sum_{c\in L}\mathbf{c}$. Avec ces définitions, on peut généraliser une propriété des codes de mots (*cf.* [4], prop. 5.2.13) : *Un ensemble de chemins C est un code si et seulement si l'on a* $(I-\mathbf{C})^{-1}=\mathbf{C}^*$ (nous notons I la matrice identité de taille S, i.e. la série caractéristique de l'ensemble des chemins vides). Notons qu'on a, classiquement : $(I-\mathbf{C})^{-1}=\sum_{n\geq 0}\mathbf{C}^n$. Un cas particulier de ceci est que, puisque l'ensemble A des arêtes est clairement un code, la série caractéristique de tous les chemins ($=\mathscr{C}$) est égale à $(I-\mathbf{A})^{-1}$.

Il n'est pas difficile de voir qu'une sous-catégorie $\mathscr{C}'$ de $\mathscr{C}$ admet un système générateur minimal : c'est l'ensemble de tous les chemins non vides dans $\mathscr{C}'$ qui ne sont pas produit de deux chemins non vides dans $\mathscr{C}'$. Par suite, si C est un code, C est entièrement défini par C^*. Ceci nous amène à introduire la définition suivante : une sous-catégorie $\mathscr{C}'$ de $\mathscr{C}$ est *libérable* si l'on a : pour tous chemins c_1, c_2, c_3, les relations c_1, c_2, c_1c_3, $c_3c_2\in\mathscr{C}'$ impliquent $c_3\in\mathscr{C}'$. Comme pour les codes de mots, on montre *qu'une sous-catégorie $\mathscr{C}'$ de $\mathscr{C}$ est libre si et seulement si $\mathscr{C}'$ est libérable.* Une conséquence immédiate de ceci est que l'*intersection d'une famille quelconque de sous-catégories libre est encore libre.*

Considérons l'exemple suivant, qui nous servira plus loin : soit S' un ensemble de sommets du graphe et $\mathscr{C}'$ défini par : $c\in\mathscr{C}'$ si et seulement si, soit c est un chemin vide, soit les deux extrémités de c sont dans S'. Il est

clair que $\mathscr{C}'$ est une sous-catégorie, dont on vérifie aisément qu'elle est libérable : $\mathscr{C}'$ est donc libre, donc $\mathscr{C}' = C^*$ pour un code C. Supposons maintenant que S' rencontre l'ensemble des sommets de tout *circuit* (un circuit est un chemin fermé sans point double) : alors C est un code fini; en effet, tout chemin c assez long dans $\mathscr{C}'$ s'écrit $c = xyz$ où y est un circuit. Alors $y = y_1 y_2$ où l'extrémité terminale de y_1 (= l'extrémité initiale de y_2) est dans S', donc $xy_1, y_2 z \in \mathscr{C}'$.

Ceci implique que C est fini, puisque un chemin dans C n'est pas produit de deux chemins non vides de $\mathscr{C}'$, et ne saurait donc être trop long.

4. *Codes complets, codes maximaux, codes complets minimaux*

Nous ferons l'hypothèse, dans ce paragraphe, que le graphe G est *fortement connexe* (f. c.), i. e. pour tous sommets s, t il existe un chemin $s \to t$.

Un chemin c est *facteur* d'un chemin c' s'il existe des chemins x, y tels que $c' = xcy$. Nous disons qu'un code C est *complet* si tout chemin c est facteur de quelque chemin dans C^*. Un résultat de base en théorie des codes de mots (Schützenberger) est qu'un code fini est complet si et seulement s'il est maximal (pour l'inclusion dans l'ensemble des codes). Ceci ne se généralise pas tel quel aux codes de chemins et nous allons le discuter maintenant.

Comme pour les codes de mots, on a que *tout code maximal est complet*. En effet, la démonstration pour les codes de mots (*cf.* [4], prop. 5.3.1) s'étend facilement, une fois faites les remarques suivantes : si G est réduit à un seul circuit, alors tout code C tel que C^* soit infini contient un chemin fermé (qui consiste à tourner plusieurs fois autour du circuit), donc C est complet; et si C^* n'est pas infini, considérons le graphe $G' = (S, A')$ où $a : s \to t$ est une arête dans G' s'il existe dans C^* un chemin non vide $s \to t$; ce graphe est sans circuits, puisque C^* est fini : soit s un sommet initial dans G' et t un sommet terminal (i. e. il n'y a pas dans C^* un chemin arrivant en s, ni de chemin partant de t); soit de plus c un chemin $s \to t$ qui ne soit pas dans C^* (il existe puisque C^* est fini). Alors on vérifie aisément que $C \cup c$ est un code. Donc C n'est pas maximal. Maintenant, si G n'est pas réduit à un circuit, comme G est fortement connexe, il y a dans G deux circuits distincts c et c' : on peut supposer que la première arête dans c n'apparaît pas dans c' et que la première arête de c' n'apparaît pas dans c. Alors pour tout chemin x tel que $c^i x c'^j$

soit défini et que $|c^i| = |c'^j| > |x|$, le chemin $c^i x c'^j$ est *sans bord*, i.e. il n'a pas de facteur gauche qui soit aussi facteur droit : on peut conclure la preuve, en la calquant sur celle pour les codes de mots.

Pour la réciproque, nous introduisons une condition sur G :

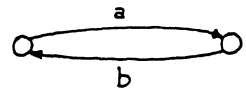
(0) pour tout sommet s, il y a une arête $s \to s$.

THÉORÈME 1. — *Soit G un graphe fortement connexe. Les conditions suivantes sont équivalentes :*

(i) *Tout code fini complet est maximal.*

(ii) *G possède la propriété* (0).

La partie (i) $\Rightarrow$ (ii) de l'énoncé est illustrée par le graphe (n'ayant pas la propriété (0))



et les deux codes $\{ab\}$, $\{ab, ba\}$. Le premier est complet, mais non maximal puisque contenu dans le second.

Preuve. — (i) $\Rightarrow$ (ii) Si G ne vérifie pas (0), soit s un sommet sans arête $s \to s$. Soit $S' = S \setminus s$. Alors S' rencontre tout circuit dans G. Considérons alors le code fini C de l'exemple à la fin du paragraphe 3. Il est clairement complet puisque G est fortement connexe; soit $c : s \to s$ un chemin non vide quelconque. Alors $C \cup c$ est un code, puisque dans C^* aucun chemin non vide ne commence ni aboutit dans s : contradiction.

(ii) $\Rightarrow$ (i) Nous avons besoin d'un résultat sur les catégories finies fortement connexes (i.e. d'un objet à un autre de la catégorie, il y a toujours un morphisme). Nous appellerons *idéal* d'une catégorie un ensemble de morphismes qui est stable par multiplication à gauche ou à droite par tout morphisme. Nous définissons de manière analogue les idéaux à gauche et à droite. Dans une catégorie finie f.c., il est clair qu'il existe un idéal qui est contenu dans tous les autres : son *idéal minimal.*

Notons aussi que si D (resp. G) est un idéal à droite (resp. à gauche) minimal de la catégorie, alors il existe un objet (nous dirons aussi sommet) s (resp. un objet t) tel que tout morphisme f dans D (resp. dans G) soit de la forme $f : s \to s'$ (resp. $f : t' \to t$) : s sera appelé *l'origine de D* (resp. l'*extrémité* de G).

THÉORÈME (de Suschkewitsch pour les catégories finies f.c.). — *Étant donnée une catégorie finie f.c. $\mathscr{D}$, son idéal minimal J est réunion disjointe*

des idéaux à droite (resp. à gauche) minimaux. Si D (resp. G) est un idéal à droite (resp. à gauche) minimal d'origine s (resp. d'extrémité t), alors $D \cap G$ est un groupe, si $s = t$. *Tous les groupes ainsi obtenus sont isomorphes.*

Ce théorème s'obtient comme conséquence du théorème de Rees-Suschkewitsch [4], (th. 3.2.7) : il suffit de considérer le semi-groupe S obtenu en ajoutant un zéro à $\mathscr{D}$, en convenant que le produit dans S de deux morphismes dont le produit dans $\mathscr{D}$ n'est pas défini est égal à 0.

Nous définissons maintenant, pour tout ensemble L de chemins dans G, la *catégorie syntaxique* de L : pour cela, nous définissons la congruence syntaxique $\sim$ de L; c'est une relation d'équivalence sur l'ensemble des chemins dans G qui est compatible avec la composition des chemins. Par définition, on aura $c \sim c'$ si et seulement si c et c' ont même origine et même extrémité et si pour tous chemins c_1, c_2, on a :

$$c_1 c c_2 \in L \quad \Leftrightarrow \quad c_1 c' c_2 \in L.$$

La catégorie syntaxique de L est définie comme la catégorie $\mathscr{D} = (S, \mathscr{C}/\sim)$. Si C est un code fini, il n'est pas difficile de montrer que la catégorie syntaxique de $L = C^*$ est finie; ceci est en fait un cas particulier de l'extension aux chemins dans un graphe du théorème de Kleene pour les mots : *un ensemble de chemins L est rationnel* (i.e. s'obtient à partir des ensembles finis par un nombre fini des trois opérations : $(L_1, L_2) \mapsto L_1 \cup L_2$, $(L_1, L_2) \mapsto L_1 L_2$, $L_1 \mapsto L_1^*$) *si et seulement si la catégorie syntaxique de L est finie.* Ce théorème peut en fait s'obtenir directement à partir du théorème de Kleene classique, comme l'avait remarqué J. BETRÉMA [3].

Nous pouvons finir la preuve du théorème. Soit C un code fini et complet, $\mathscr{D}$ la catégorie syntaxique de C^* et $\mu : \mathscr{C} \to \mathscr{D}$ le foncteur qui est l'identité sur S et associe à tout chemin sa classe mod. $\sim$.

Soit c un chemin n'appartenant pas à C. Nous montrons que $C \cup c$ n'est pas un code (donc C est maximal). Remarquons d'abord que si $x : s \to t$ est un chemin et a, b des arêtes $a : s \to s$, $b : t \to t$, comme $a^n x b^n$ peut se compléter en un chemin dans C^*, il existe i, j tels que $a^i x b^j \in C^*$ (prendre n plus grand que la longueur des chemins dans C). En particulier, pour tous états s, t, il existe un chemin $s \to t$ dans C^*.

Soient s, t l'origine et l'extrémité de c; il existe un chemin $x : t \to s$ dans C^*. Alors cx est un chemin $s \to s$.

Soit f un morphisme dans l'idéal minimal J de $\mathscr{D}$; on a $f=\mu(y)$ pour un chemin y. Comme C est complet, il existe des chemins u, v tels que $uyv\in C^*$. De plus, d'après une remarque précédente, il existe dans C^* des chemins $p : s \to \text{origine}(u)$, q : extrémité $(v) \to s$. Alors $w=puyvq : s \to s$ est dans C^* et $\mu(w)=\mu(pu)f\mu(vq)\in J$, puisque $f\in J$. Soit D (resp. G) l'idéal minimal à droite (resp. à gauche) contenant $\mu(w)$. Alors $\mu(w)$ appartient au groupe fini $D\cap G$; soit n la cardinalité de celui-ci. Alors, comme $\mu(wcxw)$ est aussi dans $D\cap G$, on a $\mu(w)^n=\mu(wcxw)^n$, donc $w^n\sim(wcxw)^n$.

Par définition de $\sim$ et comme $w^n\in C^*$, on en déduit $(wcxw)^n\in C^*$. On a donc une relation

$$wcxw\,.\,wcxw\ldots wcxw=c_1c_2\ldots c_p\,(c_i\in C),$$

ce qui montre que $C\cup c$ n'est pas un code, puisque w, $x\in C^*$ et que c apparaît dans le membre gauche et pas à droite. □

Nous dirons que C est un code *complet minimal* si C est un code complet et si pour tout code complet C' inclus dans C, on a $C'=C$. Nous notons $G(C)$ le graphe obtenu en ne conservant que les sommets intervenant comme origine ou extrémité d'un chemin dans C, et dont les arêtes sont les éléments de C.

Le fait que $G(C)$ ne soit pas fortement connexe exprime que C est composé de plusieurs morceaux qui n'ont rien à voir entre eux (*cf.* l'exemple après le théorème 1). Ceci est élucidé par le résultat suivant.

THÉORÈME 2. — *Soit* G *un graphe fortement connexe et* C *un code complet. Les deux conditions suivantes sont équivalentes :*

(i) *C est un code complet minimal.*

(ii) *G (C) est fortement connexe.*

Preuve. — Nous montrons que si C est un code complet, alors il contient un code complet C' tel que $G(C')$ soit fortement connexe (d'où (i) ⇒ (ii)). Pour cela, nous appelons *chemin infini* une application $x : \mathbb{Z} \to A$ telle que pour tout $i\in\mathbb{Z}$, le produit des arêtes $x(i)x(i+1)$ soit défini dans G. Un tel chemin infini x possède au moins un *décodage* dans C, i. e. une application strictement croissante $d : \mathbb{Z} \to \mathbb{Z}$ telle que pour tout i, $x(d(i))x(d(i)+1)\ldots x(d(i+1)-1)$ soit dans C : ceci résulte de ce que C est complet, fini et d'un argument de compacité.

Nous choisissons un chemin infini x tel que tout chemin dans G soit de la forme $x(i)\dots x(i+h)$, h, $i\geqslant 0$: c'est possible puisque G est f. c. Soit d un décodage de x dans C et $i_0\geqslant 0$ tel que le sommet initial de $x(d(i_0))$ soit sommet initial de $x(d(i))$ pour une infinité de $i\geqslant 0$. Soit C' l'ensemble des chemins de la forme

$$x(d(i))\dots x(d(i+1)-1), \qquad i\geqslant i_0.$$

Alors C' est un code contenu dans C, $G(C')$ est fortement connexe et C' est complet, par le choix de d.

Nous supposons maintenant que C est un code tel que $G(C)$ soit f. c. et que $c\in C$ est un chemin tel que $C\setminus c$ soit complet : nous en déduirons une contradiction (d'où (ii) $\rightarrow$ (i)). Soit $c : s\rightarrow t$. Comme $C\setminus c$ est complet il existe un code complet $C'\subset C\setminus c$ et tel que $G(C')$ soit f. c. (*cf.* le début de la preuve). Soient $\mathscr{D}$ la catégorie syntaxique de C'^* et $\mu : \mathscr{C}\rightarrow\mathscr{D}$ le foncteur canonique. Soit f un morphisme dans l'idéal minimal J de $\mathscr{D}$ (J existe puisque $\mathscr{D}$ est finie, *cf.* la preuve du théorème 1) et y un chemin tel que $f=\mu(y)$. Comme C' est complet, y se prolonge en un chemin dans C'^*, et comme $G(C')$ est f. c., celui-ci se prolonge en un chemin fermé et dans C'^* $z : u\rightarrow u$. Alors $\mu(z)$ est dans J donc $\mu(z)\mathscr{D}\mu(z)$ est un groupe fini G (théorème de Suschkewitsch), dont l'élément neutre e est égal à une puissance de z, donc $e\in\mu(C'^*)$. Comme $G(C)$ est f. c., il existe dans C^* des chemins $x_1 : u\rightarrow s$ et $x_2 : t\rightarrow u$. Alors $\mu(z\,x_1\,cx_2\,z)$ est dans G, donc pour $n=|G|$, $\mu(zx_1\,cx_2\,z)^n=e$. On en déduit que $(zx_1\,cx_2\,z)^n\in C'^*$, donc une relation $(zx_1\,cx_2\,z)^n=c_1\,c_2\dots c_p$, $c_i\in C'$. Mais ceci contredit le fait que C soit un code, puisque c apparaît à gauche et pas à droite. □

5. Factorisation

Nous appelons polynôme d'un code C l'élément de l'algèbre du graphe égal à $I-\mathbf{C}$; dans le cas où $C=A$, nous parlerons simplement du *polynôme du graphe*. Dans l'exemple du paragraphe 3, les polynômes du graphe et du code sont respectivement

$$I-\mathbf{A}=\begin{pmatrix}1-a & -b\\ -c & 1-d\end{pmatrix};$$

$$I-\mathbf{C}=\begin{pmatrix}1-aa-bc-abc-bdc & -ab-bd-abd-bdd\\ -dc & 1-dd\end{pmatrix}.$$

Le *déterminant du code* (fini) C est le déterminant de la matrice image canonique dans $\mathbb{Z}[A]$ (polynômes commutatifs) de $I-C$. Dans l'exemple le *déterminant du graphe* (i. e. du code A) est $1-a-d+ad-bc$ et le déterminant de C est $1-a^2-bc-d^2-abc-bcd+a^2d^2-abcd$. Nous utiliserons le résultat suivant, intéressant en soi.

THÉORÈME 3. — *Si G est un graphe fortement connexe, alors sont déterminant est un polynôme irréductible.*

Remarques. — 1. La réciproque est *presque* vraie : le déterminant de G est en effet égal au produit des déterminants de ses composantes fortement connexes; donc, *si le graphe a au moins deux composantes fortement connexes non triviales* (graphe trivial = graphe sans arêtes), *alors son déterminant est réductible.*

2. Notons $\bar{c}$ le monôme dans $\mathbb{Z}[A]$ égal au produit des arêtes (avec leurs multiplicités) intervenant dans le chemin c du graphe $G=(S, A)$. Un calcul classique montre qu'alors le déterminant D de G est

$$D=1+\sum_{k\geqslant 1}(-1)^k\sum_{c_1,\ldots,c_k}\bar{c}_1\ldots\bar{c}_k,$$

où la deuxième sommation est étendue à tous les circuits (= chemins fermés simples) $c_1, \ldots, c_k$ deux à deux disjoints (i. e. sans sommets communs). D est donc de degré partiel $\leqslant 1$ en chacune des arêtes. On montre facilement que tous les termes du développement ci-dessus sont distincts. En particulier, D est à coefficients 1 et -1 et, si G est fortement connexe, le degré partiel de D est 1 en chaque arête (puisque chaque arête fait partie d'un circuit). Une autre remarque sur les circuits est que si c, c' sont deux circuits tels que $\bar{c}$ divise $\bar{c}'$, alors $c=c'$ et $\bar{c}=\bar{c}'$.

3. Soit D le déterminant d'un graphe fortement connexe. La connaissance de D implique celle de l'ensemble des arêtes de chaque circuit, ainsi que des couples de circuits disjoints; elle ne suffit pas à caractériser le graphe. Mais THOMASSEN [10] a montré que le graphe est déterminé par la donnée des ensembles d'arêtes de ses circuits à quelques transformations simples près : il s'agit d'une extension aux graphes orientés d'un théorème de Whitney (*voir* [11] pour une preuve simple de ce dernier).

Preuve. — Soit D le déterminant de G. Écrivons que $D=PQ$, avec P, Q dans $\mathbb{Z}[A]$, de termes constants 1. Supposons, par l'absurde, que dans P ou Q apparaisse un monôme $x\neq 1$, diviseur strict d'un $\bar{c}$ (c circuit) : choisissons $|x|$ minimum et disons que x apparaît dans P. Alors

$(D, x)=(P, x)+\sum_{uv=x,\, v\neq 1}(P, u)(Q, v)$(où (P, u) désigne le coefficient de u dans P). Par minimalité de x, on ne peut avoir $(P, u)\,(Q, v)\neq 0$ que si $u=1$, $v=x$; mais dans ce cas, le degré de Q en a, a variable dans x, est $\geqslant 1$, ce qui est déjà le cas de P : contradiction avec la remarque 2 *ci-dessus*. *Donc* $(D, x)=(P, x)\neq 0$: mais ceci n'est pas possible, car on aurait alors $x=\bar{c}_1 \ldots \bar{c}_k (k\geqslant 1, c_1, \ldots, c_k$ circuits dans $G)$, d'où : $\bar{c}_1$ diviseur strict de $\bar{c}$, en contradiction avec la remarque 2.

Pour tout circuit c, on a $0\neq(D, \bar{c})=\sum_{\bar{c}=uv}(P, u)(Q, v)$; ce qui précède montre alors que $\bar{c}$ apparaît dans P ou Q. Il ne peut apparaître dans les deux, d'après la même remarque. On a donc une réunion disjointe : $C=C_1 \cup C_2$ (où C est l'ensemble des circuits de G) où C_1 (resp. C_2) est l'ensemble des circuits dont l'image commutative apparaît dans P (resp. Q). Soient $c_1\in C_1$, $c_2\in C_2$. Alors

$$(D, \bar{c}_1 \bar{c}_2)=\sum_{uv=\bar{c}_1 \bar{c}_2}(P, u)(Q, v)$$

ne peut-être nul que s'il existe une factorisation

$$uv=\bar{c}_1 \bar{c}_2, \qquad (u, v)\neq(\bar{c}_1, \bar{c}_2), \qquad (P, u)(Q, v)\neq 0 :$$

i.e. $\bar{c}_1=xy$, $\bar{c}_2=zt$, $u=xz$, $v=yt$ et $y\neq 1$ ou $z\neq 1$. Si $y\neq 1$, soit a une variable dans y; alors a apparaît dans P (par $\bar{c}_1$) et dans Q (par v) : contradiction. Il en est de même si $z\neq 1$. On a donc : $(D, \bar{c}_1 \bar{c}_2)\neq 0$.

Ce qui précède montre que pour tous $c_1\in C_1$, $c_2\in C_2$, les deux circuits c_1 et c_2 sont disjoints (puisque $\bar{c}_1 \bar{c}_2$ apparaît dans D). Si G est fortement connexe, on doit donc avoir $C_1=\varnothing$ ou $C_2=\varnothing$: en effet, soit $S_i (i=1, 2)$ l'ensemble des sommets des circuits dans C_i, et soit $s_1\in S_1$, $s_2\in S_2$; il existe un chemin $s_1 \to s_2$, donc il existe une arête $a : s\to t$ avec $s\in S_1$, $t\in S_2$. Cette arête fait partie d'un circuit c; s se trouve sur un circuit $c_1\in C_1$ et t sur $c_2\in C_2$: mais alors, soit $c\in C_1$ et c et c_2 ne sont pas disjoints, soit $c\in C_2$ et c et c_1 ne sont pas disjoints : contradiction.

Supposons $C_2=\varnothing$. Alors tout circuit apparaît dans P, donc toute variable $a\in A$ aussi. On a donc $Q=1$, sinon il existerait une variable apparaissant à la fois dans P et Q, en contradiction avec la remarque 2. □

Nous démontrons le :

THÉORÈME 4. — *Soit G un graphe fortement connexe vérifiant l'hypothèse* (0). *Alors le déterminant de tout code fini complet est divisible par le déterminant du graphe G.*

Dans l'exemple ci-dessus, on a

$$1-a^2-bc-d^2-abc-bcd+a^2d^2-abcd$$
$$=(1-a-d+ad-bc)(1+a+d+ad).$$

Dans le théorème 2, l'hypothèse « fortement connexe » n'est pas nécessaire; en effet, dans le cas général, on peut mettre les matrices sous forme triangulaire par blocs : chaque bloc correspondant à une composante fortement connexe et l'ensemble général s'en déduit, moyennant le théorème 4. Mais cet énoncé plus général n'est pas vraiment pertinent, *cf.* la discussion du paragraphe 6.

Preuve. — Nous reprenons les notations et définitions de la partie (ii) $\Rightarrow$ (i) de la preuve du théorème 1. Soit $\mathscr{D}$ la catégorie syntaxique du code fini complet C, μ le foncteur canonique $\mathscr{C} \to \mathscr{D}$, s_0 un sommet distingué du graphe G, w dans C^* un chemin $s_0 \to s_0$ tel que $e=\mu(w)$ soit dans l'idéal minimal J de $\mathscr{D}$, Δ (resp. Γ) l'idéal à droite (resp. à gauche) minimal contenant e; $H=\Delta \cap \Gamma$ est un groupe fini et on peut donc supposer, quitte à remplacer w par $w^{|H|}$, que e est l'élément neutre de H. L'ensemble $K=\{f\in H \mid \exists c\in C^*,\ \mu(c)=f\}$ est un sous-monoïde de H (puisque $\mu(c)$, $\mu(c')\in H$ implique en particulier que c, c' sont des chemins $s_0 \to s_0$), donc un sous-groupe de H. Soient $x_1, \ldots, x_d$ des chemins $s_0 \to s_0$ dans G tels que $K\mu(x_1), \ldots, K\mu(x_d)$ soient les d classes à gauche de H mod. K. Soient $y_1, \ldots, y_d$ des chemins $s_0 \to s_0$ tels que $\mu(x_i)$ soit l'inverse de $\mu(y_i)$ dans H, pour tout i. On peut supposer que $\mu(x_1)=\mu(y_1)=e$. Pour tout couple de sommets (s, t), soit $z_{s,t} : s \to t$ un chemin dans C^* (il existe d'après la preuve du théorème 1). Alors pour tout chemin c, il existe un et un seul triplet $(s, t, i)\in S\times S\times\{1, \ldots, d\}$ tel que

$$x_1 z_{s_0,s} c z_{t,s_0} y_i \in C^*.$$

En effet, si cette relation est vraie, on a forcément que c est un chemin $s \to t$ et que

$$\mu(x_1)\mu(z_{s_0,s}cz_{t,s_0})\mu(y_i)\in\Delta\cap\Gamma\cap\mu(C^*)=K,$$

donc $e\mu(z_{s_0,s}cz_{t,s_0})e\in K\mu(x_i)$, donc i est entièrement déterminé par la classe mod. K qui contient l'élément $e\mu(z_{s_0,s}cz_{t,s_0})e$ de H. L'existence du triplet se montre de manière analogue. Nous poserons :

$$u_s = x_1 z_{s_0, s}, \qquad v_{i,t} = z_{t, s_0} y_i.$$

On a donc une réunion disjointe

$$\mathscr{G} = \cup_{s,t \in S,\ 1 \leq i \leq d} E(u_s, v_{i,t}),$$

où $E(x, y) = \{ c \in \mathscr{G} \mid xcy \in C^* \}$.

Définissons, pour chaque chemin x, les ensembles

$$S'(x) = \{ c \in \mathscr{G} \mid xc \in C^* \}, \qquad S(x) = S'(x) \setminus S'(x) C,$$
$$P'(y) = \{ c \in \mathscr{G} \mid cy \in C^* \}, \qquad P(y) = P'(y) \setminus C P'(y).$$

Alors, $S(x)$ est composé de facteurs droits de mots de C, donc est fini; de même, $P(y)$ est fini. Il est facile de voir que $S(x) C^* P(y) \subset E(x, y)$. De plus,

$$F(x, y) = E(x, y) \setminus S(x) C^* P(y)$$

est composé de facteurs de mots dans C, donc est fini. Par ailleurs, le fait que C est un code implique que le produit $S(x) C^* P(y)$ est *non ambigu*, i. e. si s, $s' \in S(x)$, m, $m' \in C^*$, p, $p' \in P(y)$, alors $smp = s' m' p'$ implique $s = s'$, $m = m'$, $p = p'$.

Ceci montre que dans l'algèbre large du graphe, on a la relation en séries caractéristiques :

$$\mathscr{G} = \sum_{s,t \in S,\ 1 \leq i \leq d} (\mathbf{S}(u_s) \mathbf{C}^* \mathbf{P}(v_{i,t}) + \mathbf{F}(u_s, v_{i,t})) \tag{5.1}$$

d'où

$$\mathscr{G} = \mathbf{S} \mathbf{C}^* \mathbf{P} + \mathbf{F}$$

où l'on a posé

$$S = \sum_{s \in S} \mathbf{S}(u_s), \qquad \mathbf{P} = \sum_{t \in S,\ 1 \leq i \leq d} \mathbf{P}(v_{i,t})$$

et

$$\mathbf{F} = \sum_{s,t \in S,\ 1 \leq i \leq d} \mathbf{F}(u_s, v_{i,t}).$$

On remarquera que $\mathbf{F}$ est une combinaison linéaire (finie) de $\mathbf{c}$, pour des chemins c non vides : en effet, le chemin vide 1_s apparaît déjà dans $\mathbf{S}(u_s) \mathbf{C}^* \mathbf{P}(v_{1,t})$ puisque $u_s = x_1 z_{s_0, s}$, $v_{1,t} = z_{t, s_0} y_1$ sont dans C^* (car $\mu(x_1) = \mu(y_1) = e \in \mu(C^*)$), donc 1_s appartient à $S'(u_s)$ et à $P'(v_{1,t})$ et enfin

à $S(u_s)$ et à $P(v_{1,t})$. Comme 1_s apparaît une seule fois à gauche dans (5.1), il n'apparaît dans aucun $\mathbf{F}(u_s, v_{i,t})$, donc pas non plus dans $\mathbf{F}$.

Nous avons donc

$$\mathbf{A}^* - \mathbf{F} = (I-\mathbf{A})^{-1} - \mathbf{F} = \mathbf{S}(I-\mathbf{C})^{-1}\mathbf{P}$$

d'où

$$(I-\mathbf{A})^{-1}(I-(I-\mathbf{A})\mathbf{F}) = \mathbf{S}(I-\mathbf{C})^{-1}\mathbf{P}.$$

La matrice constante de $I-(I-\mathbf{A})\mathbf{F}$ est I (puisqu'aucun chemin vide n'apparaît dans $\mathbf{F}$), donc les matrices constantes de $\mathbf{S}$ et $\mathbf{P}$ sont inversibles. Tous ces éléments sont donc inversibles dans l'algèbre large du graphe, et l'on obtient

$$\mathbf{P}^{-1}(I-\mathbf{C})S^{-1} = (I-(I-\mathbf{A})\mathbf{F})^{-1}(I-\mathbf{A})$$

d'où

$$I-\mathbf{C} = \mathbf{P}(I-(I-\mathbf{A})\mathbf{F})^{-1}(I-\mathbf{A})\mathbf{S}.$$

Soit ρ l'homomorphisme canonique $\mathbb{Z}\langle\langle A\rangle\rangle \to \mathbb{Z}[[A]]$. On a alors

$$\rho(I-\mathbf{C}) = \rho(\mathbf{P})\,\rho(I-(I-\mathbf{A})\mathbf{F})^{-1}\,\rho(I-\mathbf{A})\,\rho(\mathbf{S})$$

et en passant aux déterminants :

$$\det(\rho(I-\mathbf{C})) = \frac{\det(\rho(\mathbf{P}))\det(\rho(I-\mathbf{A}))\det(\rho(\mathbf{S}))}{\det(\rho(I-(I-\mathbf{A})\mathbf{F}))}.$$

Comme le membre gauche est un polynôme, le dénominateur de droite se factorise en $R_1R_2R_3$ où R_1, R_2, R_3 divisent respectivement les trois facteurs du numérateur (avec $R_2 \mid \det\rho(I-\mathbf{A})$). Mais $\det\rho(I-\mathbf{A})$ est irréductible par le théorème 3 et ne peut diviser $\det\rho(I-(I-\mathbf{A})\mathbf{F})$: en effet l'hypothèse (0) implique qu'on peut trouver des valeurs des variables dans A telles que $\mathbf{A}=I$ et par suite : $\det\rho(I-\mathbf{A})=0$ mais $\det\rho(I-(I-\mathbf{A})\mathbf{F})\neq 0$. Donc $\det\rho(I-\mathbf{A})$ et $\det\rho(I-(I-\mathbf{A})\mathbf{F})$ sont premiers entre eux, d'où l'on déduit que $R_2 = \pm 1$. Ceci montre que le déterminant de C est divisible par celui du graphe, et achève la preuve. □

6. Problèmes et perspectives

Nous conjecturons que le théorème 4 est encore vrai sans l'hypothèse (0) sur le graphe. Plus généralement, nous pensons que pour tout code C

fini et complet minimal d'un graphe fortement connexe $G=(S, A)$, il existe des éléments P, S de l'algèbre du graphe tels que

$$I-\mathbf{C}=X(I-\mathbf{A})Y. \tag{6.1}$$

Pour l'exemple du paragraphe 3, on a :

$$\begin{pmatrix} 1-aa-bc-abc-bdc, & -ab-bd-abd-bdd \\ -dc, & 1-dd \end{pmatrix} = \begin{pmatrix} 1+a, & b \\ 0, & 1 \end{pmatrix} \begin{pmatrix} 1-a, & -b \\ -c, & 1-d \end{pmatrix} \begin{pmatrix} 1, & 0 \\ c, & 1+d \end{pmatrix}.$$

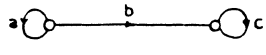
L'analogue de cette conjecture, dans les cas des codes de mots, est démontrée dans [7] et [8]. Dans tous les exemples connus, les matrices X et Y sont les matrices caractéristiques d'ensemble finis de chemins P et S. On obtient alors, par inversion dans l'algèbre large du graphe, la relation

$$\mathbf{A}^*=\mathbf{SC}^*\mathbf{P} \tag{6.2}$$

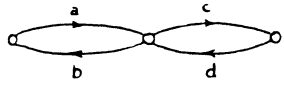
Cette relation s'exprime en disant que tout chemin c du graphe admet une et une seule décomposition de la forme

$$c=sc_1 \ldots c_n p, \qquad s\in S,\ c_i\in C,\quad n\geqslant 0,\quad p\in P$$

Cette dernière formulation de la conjecture (la plus forte) généralise la *conjecture de factoirisation* pour les codes de mots (Schützenberger, Césari, Perrin). Il faut remarquer qu'on ne peut espérer affaiblir les hypothèses. En effet, pour le graphe (non fortement connexe)



et le code $C=\{a, b, ab, c\}$, on ne peut avoir une relation des la forme (6.1). De même pour le graphe



et le code (non complet minimal)

$C=\{a, b, cd, dc\}$.

Il y a d'autres résultats de la théorie des codes de mots qui semblent s'étendre aux codes de chemins. En particulier, toute la théorie combinatoire des codes bipréfixes : ceci sera fait par ailleurs. Il serait intéressant

aussi d'examiner ce que deviennent les résultats dans [5] et [6] : double transitivité des groupes associés aux codes bipréfixes « premiers », semi-simplicité de l'algèbre associée aux codes bipréfixes. Un autre problème est de voir comment s'étend la notion de *degré* d'un code et le théorème de factorisation commutative de Schützenberger [9].

Remerciements

Je tiens à remercier M. P. Schützenberger pour les nombreuses discussions sur les codes, qui ont été à l'origine de ce travail. Et aussi C. Thomassen qui a consacré une soirée à la Rotonde à m'expliquer le théorème de Whitney sur les graphes et son extension aux graphes orientés.

RÉFÉRENCES

[1] BERSTEL (J.) and PERRIN (D.), *Theory of codes*, Acad. Press (à paraître).

[2] BETRÉMA (J.), Classification et représentation de systèmes d'actions, *Thèse 3e cycle*, Université de Paris-7, 1981.

[3] GABRIEL (P.), Auslender-Reiten sequences and representation-finite algebras, *Lecture Notes Maths.*, t. 831, 1980, p. 1-71.

[4] LALLEMENT (G.), *Semigroups and combinatorial application*, John Wiley, 1979.

[5] PERRIN (D.), *Sur la transitivité du groupe d'un code bipréfixe Math. Zeitschrift*, t. 153, 1877, p. 283-287.

[6] REUTENAUER (C.), Semisimplicity of the algebra associated to a biprefix code, *Semigroup Forum*, vol. 23, 1981, p. 327-342.

[7] REUTENAUER (C.), Sulla fattorizzazione dei codici, *Ricerche di Matematica*, vol. 32, 1983, p. 115-130.

[8] REUTENAUER (C.), Noncommutative factorization of variable-length codes, *J. Pure Applied Algebra* (à paraître).

[9] SCHÜTZENBERGER (M. P.), Sur certains sous-monoïdes libres, *Bull. Soc. Math. Fr.*, vol. 93, 1965, p. 209-223.

[10] THOMASSEN (C.), Communication personnelle.

[11] TRUEMPER (K.), *On Whitney's 2-isomorphism theorem for graphs.*

Ajouté sur épreuves : La théorie combinatoire des codes bipréfixes a été récemment étendue par C. De Felice aux codes de chemins (Actes du colloque ICALP 86, *Lecture Notes Computer Sci.*, à paraître).