

ALAIN GUÉNOCHE

**Vainqueurs de Kemeny et tournois difficiles**

*Mathématiques et sciences humaines*, tome 133 (1996), p. 57-65

[http://www.numdam.org/item?id=MSH\\_1996\\_\\_133\\_\\_57\\_0](http://www.numdam.org/item?id=MSH_1996__133__57_0)

© Centre d'analyse et de mathématiques sociales de l'EHESS, 1996, tous droits réservés.

L'accès aux archives de la revue « Mathématiques et sciences humaines » (<http://msh.revues.org/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

## VAINQUEURS DE KEMENY ET TOURNOIS DIFFICILES

Alain GUÉNOCHE<sup>1</sup>

**RÉSUMÉ** – *Dans cet article, on s'intéresse à la détermination des ordres médians des tournois valués. On propose d'une part des améliorations d'une méthode arborescente permettant de limiter le nombre de nœuds et donc d'accélérer l'énumération des ordres médians. D'autre part, pour les tournois difficiles qui restent incalculables, on propose de réduire le tournoi en éliminant certains candidats.*

**SUMMARY** – *Kemeny winners and hard tournaments*

*In this paper, we deal with the computation of median orders of weighted tournaments. First, we present improvements of a branch and bound method in order to speed up the enumeration of median orders. Then, for the hard tournaments for which these improvements are not sufficient, we study two ways to reduce the tournament by deleting vertices which appear as poor candidates.*

### 1. INTRODUCTION

Un tournoi est un graphe orienté, antisymétrique et complet. C'est par exemple le résultat de comparaisons par paires. Sur un ensemble  $X$  de candidats, par un système de notes, des classements d'experts ou des comparaisons deux à deux, on décide si  $x$  est meilleur que  $y$  et on oriente en conséquence l'arc entre  $x$  et  $y$ . Cette préférence peut être pondérée ou non, donnant ainsi un tournoi valué ou non. Pour les tournois valués, on admet des arcs de poids nul, c'est-à-dire qu'il peut y avoir indifférence entre deux candidats.

Nous nous posons le problème de désigner un sous-ensemble des candidats, les *gagnants* de ce tournoi, ceux qui sont considérés collectivement comme les meilleurs. Suivant le principe de Condorcet [1785], s'il existe un élément préféré à tous les autres, c'est lui qui doit être désigné vainqueur ; mais souvent aucun élément ne remplit cette condition. D'autre part, on ne veut souvent pas qu'un seul gagnant, mais plusieurs candidats sélectionnés, voire un ordre total sur l'ensemble des candidats.

Ce problème se rencontre dans de nombreux domaines d'application ; citons entre autres :

- le dépouillement d'un scrutin, pour élire un certain nombre de représentants ;
- le classement des élèves d'après plusieurs épreuves, pour l'attribution des bourses ;
- le choix agronomique, parmi des variétés d'une même espèce, en fonction du rendement ou d'autres critères ;
- les classements hédonistes, pour attribuer des médailles après des dégustations comparatives ;
- le choix d'une heuristique parmi plusieurs méthodes concurrentes pour des problèmes NP-difficiles d'optimisation.

---

<sup>1</sup> LIM-CNRS, 163 Av. de Luminy, 13288 Marseille Cedex 9, guenoche@lim.univ-mrs.fr

Soit  $T$  un *tournoi valué*, dont les sommets  $\{1, \dots, n\}$  correspondent aux éléments de  $X = \{x_1, \dots, x_n\}$ , et entre les sommets  $i$  et  $j$ , on a l'arc  $(i, j)$  (resp.  $(j, i)$ ) si  $x_i$  est préféré à  $x_j$  (resp.  $x_j$  à  $x_i$ ). Cet arc possède une valeur entière positive ou nulle notée  $T(i, j)$  proportionnelle au degré de préférence en faveur de  $x_i$ . Si  $T(i, j) > 0$ , on a  $T(j, i) = 0$ . Si les éléments  $x_i$  et  $x_j$  sont considérés comme équivalents, on a  $T(i, j) = T(j, i) = 0$ .

Pour désigner les gagnants, on énumère les ordres totaux à distance minimum de ce tournoi, encore appelés *ordres médians* [Barthélemy et Monjardet 1981]. Si le tournoi  $T$  est transitif, il lui correspond un ordre total unique compatible avec lui. Sinon, pour tout ordre  $O = o_1 > o_2 > \dots > o_n$ , il y a dans  $T$  au moins un arc  $(o_j, o_i)$  avec  $j > i$  par lequel passe un circuit ; c'est l'*effet Condorcet*. Classiquement on place les sommets sur un axe, de gauche à droite en suivant cet ordre et on ne trace que les arcs du tournoi qui sont en désaccord avec l'ordre, et qui vont donc de droite à gauche et que l'on appelle *arcs retour* [Hudry 1989]. Ainsi, la distance entre le tournoi  $T$  et un ordre  $O$  est la somme des poids des arcs retour :

$$d_K(O, T) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n T(o_j, o_i)$$

Pour rendre le tournoi transitif, il faut supprimer tous ses circuits, c'est-à-dire inverser certains arcs. Si le tournoi est non valué, tous les arcs ont un même poids unitaire et on cherche à inverser un nombre minimum d'arcs, mais s'il est valué, il convient d'inverser les arcs dont la somme des poids est minimum. Les ordres totaux à distance minimum (les *ordres médians*) du tournoi, sont ceux qui minimisent  $d_K(O, T)$ .

Soulignons d'emblée qu'il y a rarement un seul ordre à distance minimum. Il peut y avoir beaucoup d'*ex æquo*, et même une quantité exponentielle [Hudry et Woïrgard 1994]. Ils sont tous équivalents, c'est pourquoi il est nécessaire de les énumérer. Les gagnants que nous désignons sont, suivant Kemeny [1959] (d'où ici la notation  $d_K$  pour la distance considérée), le ou les premiers éléments de ces ordres médians.

## 2. ÉNUMÉRATION DES ORDRES MÉDIANS

L'algorithme de base est une méthode *branch and bound* [Guénoche 1977] précédée d'une approximation de la distance minimum d'un ordre au tournoi par des méthodes heuristiques. Cette méthode est tout à fait fondée, puisque le problème est NP-difficile [voir par exemple Charon, Hudry et Woïrgard 1996] ; elle a été décrite en détail dans Barthélemy, Guénoche et Hudry [1989], et nous ne la reprendrons que dans ses grandes lignes.

On calcule préalablement la distance de certains ordres, dont l'expérience montre qu'ils sont souvent proches du tournoi. Ils peuvent être construits par des heuristiques fondées sur l'élimination des circuits de longueur 3 [Smith et Payne 1974] adaptées aux tournois valués ou, plus récemment, par des méthodes d'optimisation stochastique. Le meilleur d'entre eux fournit une borne supérieure  $D_{max}$  qu'il est inutile de dépasser.

Comme pour toute méthode *branch and bound*, on construit une arborescence dont les nœuds sont ici des débuts d'ordres totaux, des *sections commençantes*. Chacune de ces sections peut, *a priori*, se prolonger en un ordre total à distance minimum du tournoi. Chaque section a une valeur qui est une borne inférieure de la distance au tournoi de tout ordre commençant par cette section. À chaque étape, on prolonge la section de valeur minimum de toutes les manières possibles, si les sections résultantes restent à une distance inférieure ou égale à  $D_{max}$ .

Soit  $O = o_1 > \dots > o_p$  une feuille de cette arborescence, c'est-à-dire un ordre total sur les  $p$  sommets ainsi classés. Sa valeur  $V(O)$  est définie par :

$$\sum_{1 \leq i < j \leq p} T(o_j, o_i) + \sum_{i=1}^p \left[ \sum_{x \notin \{o_1, \dots, o_p\}} T(x, o_i) \right]$$

(toujours avec  $T(x, y) = 0$  si  $(x, y)$  n'est pas un arc de  $T$ ).

C'est la somme des poids des arcs retour dont l'extrémité est dans la section  $O$ , où que soit l'origine. Ses successeurs possibles sont les sections obtenues en prolongeant  $O$  avec l'un des  $n - p$  sommets  $\{y_1, \dots, y_{n-p}\}$  non placés dans  $O$ . Tant que l'on n'a pas construit un ordre total,

on répète les opérations (i) et (ii) :

- (i) Trouver une feuille  $O = o_1 > \dots > o_p$  dans l'arborescence dont la valeur est minimum. On développe donc une arborescence « au moindre coût » d'abord.
- (ii) Si  $p = n - 1$ , il suffit de lui ajouter le dernier sommet non placé  $y_1$  et l'ordre  $o_1 > \dots > o_{n-1} > y_1$  est à distance minimum du tournoi. On a donc trouvé un ordre solution à distance  $\Delta = V(O)$ .

Sinon on développe l'arborescence ; on ajoute à  $O$  (qui n'est donc plus une feuille) les nœuds correspondant à cette section prolongée de certains des  $n - p$  sommets non placés dans  $X$ . Ce sont ceux qui :

- d'une part ne dominent pas  $o_p$  si les valuations sont strictement positives, puisqu'alors un ordre à distance minimum d'un tournoi est associé à un chemin hamiltonien de ce tournoi [Remage Jr. et Thompson Jr. 1966],
- d'autre part ont une valeur inférieure ou égale à la borne  $D_{max}$  calculée initialement.

La valeur de la section  $O$  prolongée d'un sommet  $y_i$  est égale à la valeur de  $O$  augmentée du coût de  $y_i$ , c'est-à-dire la somme des valeurs des arcs allant d'un sommet  $y_k$  non classé à  $y_i$  :

$$V(o_1 > o_2 > \dots > o_p > y_i) = V(O) + \sum_{k \neq i} T(y_k, y_i)$$

Si l'on cherche une seule solution, on s'arrêtera au premier ordre total construit, mais si l'on veut tous les ordres médians, on continuera à développer l'arborescence. Si l'on peut encore prolonger quelques sections commençantes sans dépasser  $\Delta$ , on aboutit à d'autres solutions. Quand toutes les feuilles de l'arborescence correspondent à des sections impossibles à prolonger, on a énuméré tous les ordres totaux à distance minimum du tournoi.

Exemple : Soit le tournoi non valué donné par la table ci-dessous :

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>
<i>a</i>	-	1	1	0	1	1	1
<i>b</i>	0	-	1	1	0	1	1
<i>c</i>	0	0	-	1	1	0	1
<i>d</i>	1	0	0	-	1	0	0
<i>e</i>	0	1	0	0	-	1	0
<i>f</i>	0	0	1	1	0	-	0
<i>g</i>	0	0	0	1	1	1	-

On obtient aisément un ordre à distance 4,  $a > b > c > g > d > e > f$ . Cette borne permet d'énumérer 5 ordres médians, qui sont bien à distance 4 du tournoi. En plus du précédent, il s'agit de  $a > b > c > g > e > f > d$ ,  $a > b > c > g > f > d > e$ ,  $a > b > f > c > g > d > e$  et  $a > b > g > f > c > d > e$ .

Souvent ce type de méthodes échoue à cause de la complexité des problèmes de consensus dans les relations d'ordre [Barthélemy et Monjardet 1981]. La construction des ordres médians est un problème NP-difficile et l'espace mémoire nécessaire est proportionnel à la taille de l'arborescence. Même pour des programmes efficaces, il se peut que sur un tournoi d'une vingtaine de candidats on ne puisse énumérer les solutions, ni même en construire une seule. C'est à ce type de difficultés que nous consacrons le paragraphe suivant.

### 3. COMMENT LIMITER LA TAILLE DE L'ARBORESCENCE ?

C'est bien le paramètre essentiel de la calculabilité effective de l'énumération. Plus l'arborescence de recherche est grande, plus l'on passe de temps à explorer tous ses nœuds pour déterminer s'ils peuvent aboutir à un ordre médian. Voyons deux idées pour éviter de développer des nœuds inutiles.

### 3.1. Le meilleur ordre sur toute partie

Un nœud de l'arborescence est un ordre total sur un sous-ensemble de candidats classés. Mais cette partie se retrouve chaque fois que l'on aboutit au même sous-ensemble dans un ordre différent. L'idée est de conserver, pour chaque partie rencontrée, la plus petite valeur trouvée sur cette partie. Si par exemple on trouve que la section commençante  $1 > 2 > 3 > 4$  est de valeur  $v$ , on gardera cette valeur pour la partie  $\{1, 2, 3, 4\}$ . Si par la suite on aboutit à la section  $2 > 1 > 4 > 3$  de valeur  $v'$ , de trois choses l'une :

- i) si  $v' = v$ , on conserve cette section commençante dans l'arborescence si on veut tous les ordres médians ;
- ii) si  $v' < v$ , on conserve cette section commençante, on met à jour la valeur de la partie :  $V(\{1, 2, 3, 4\}) = v'$  et on annule dans l'arborescence l'ancienne section commençante  $1 > 2 > 3 > 4$ , puisque
- iii) si  $v' > v$ , il est inutile de créer le nœud correspondant à  $2 > 1 > 4 > 3$  car cette section commençante ne peut aboutir à un ordre médian.

En testant l'efficacité de cette idée pour l'énumération des ordres médians, on constate que l'arborescence résultante a, en gros, trois fois moins de sommets. Il faut certes adapter une structure de données, de type fonction caractéristique des parties de  $X$ , pour garder la valeur du meilleur ordre de chaque partie. Mais au total on est largement gagnant en temps et en place.

Un autre intérêt de cette implémentation est que si l'on ne cherche qu'un seul ordre médian, on peut confondre les cas i) et iii). Si l'on a déjà un ordre de valeur  $v$  sur cette partie, il est inutile d'encombrer l'arborescence avec un nouvel ordre de même valeur sur cette même partie. Nous avons insisté, pour notre problème, sur la nécessité d'avoir tous les ordres médians, ou tout au moins tous les « gagnants » d'un ordre médian. Ici, l'obtention d'un unique ordre médian aura pour effet de donner à  $D_{max}$  sa plus petite valeur  $\Delta$  donc, en relançant le programme d'énumération, d'aboutir, peut-être, à l'énumération complète.

### 3.2. Une valeur plus ajustée pour les sections commençantes

Pour évaluer une section commençante comme une borne inférieure de la distance au tournoi d'un ordre qui commence par cette section, dans la procédure présentée au §2 on n'utilise que les arcs retour dont l'extrémité est dans cette partie. Nous allons aussi tenir compte du poids de certains arcs dont l'origine et l'extrémité sont dans la partie complémentaire notée  $F$ .

Pour les tournois non valués, Charon-Fournier, Germa et Hudry [1992] ont utilisé la séquence des demi-degrés extérieurs du tournoi réduit à  $F$ . Un tournoi transitif d'ordre  $q$  a comme séquence de degrés, rangés dans l'ordre décroissant, la suite  $\partial_1 = q - 1, \partial_2 = q - 2, \dots, \partial_q = 0$ . Sur l'ensemble des sommets non classés on a, toujours suivant l'ordre décroissant, une séquence particulière  $d_1, \dots, d_q$ . L'inversion d'un arc abaisse le degré de son origine d'une unité, et augmente d'autant le degré de son extrémité. Donc pour passer de la séquence  $d_1, \dots, d_q$  à la séquence  $\partial_1, \dots, \partial_q$ , il faudra inverser au minimum un nombre d'arcs

$$\sigma(F) = \frac{1}{2} \sum_{i=1}^q |d_i - \partial_i|.$$

On peut par conséquent donner à une section commençante une valeur plus forte en ajoutant au nombre d'arcs retour dont l'extrémité est dans  $O$  cette valeur  $\sigma(F)$ . Les nœuds dont la valeur outrepassa  $D_{max}$  étant éliminés, l'arborescence développée ne peut être que plus petite. Expérimentalement on constate que le gain est très important, puisque le nombre de sommets de l'arborescence est diminué en moyenne de plus de 75 %. Nous avons repris cette idée en considérant les circuits de longueur 3 ; pour les tournois valués et non valués, nous avons exploité deux idées qui se sont avérées très efficaces.

### 3.2.1 Cas des tournois valués

Pour évaluer chaque section commençante  $O$ , on ne peut calculer un ordre optimal sur son complémentaire  $F$ , mais on peut minorer la part de distance due à un ordre sur  $F$  à l'aide de ses seuls circuits de longueur 3. Pour tout 3-circuit  $(x, y, z)$ , le sous-tournoi s'écrit :

	$x$	$y$	$z$
$x$	-	$a$	0
$y$	0	-	$b$
$z$	$c$	0	-

Ce circuit coûtera au moins  $v'(x, y, z) = \text{Min}\{a, b, c\}$ . On va donc considérer le plus possible de 3-circuits inclus dans  $F$ . En effet, si l'un des sommets est dans  $O$ , il se peut que l'arc de poids minimum ait pour extrémité ce sommet et cet arc est déjà comptabilisé dans  $V(O)$ . Ensuite, si l'on tient compte du circuit  $(x, y, z)$ , on ne peut considérer un autre circuit qui aurait un arc commun avec  $(x, y, z)$  car en retournant celui-là on annulerait deux circuits d'un seul coup.

Il faut donc que les circuits considérés soient arc-disjoints et de somme des poids aussi grande que possible. Pour être efficace, on utilise un algorithme glouton. Dans un pré-traitement on détermine le coût de tous les circuits de longueur 3, ce qui se fait en  $O(n^3)$ . Préalablement au développement de l'arborescence, on range les 3-circuits dans l'ordre des coûts décroissants.

Pour une section finissante  $F$ , on définit  $V'(F)$  comme la somme des coûts des 3-circuits arc-disjoints contenus dans  $F$  et sélectionnés par l'algorithme glouton. On peut alors utiliser comme valuation d'une section commençante :

$$V''(O) = V(O) + V'(F)$$

Cette meilleure borne permet de développer une arborescence très restreinte, puisque sur des tournois construits à partir d'ordres tirés au hasard et valués par le nombre de préférences, les arborescences résultantes ont 20 fois moins de sommets ! De plus cette évaluation des 3-circuits n'est pas très pénalisante du point de vue temps de calcul, puisqu'elle est en général proportionnelle au nombre de 3-circuits.

### 3.2.2 Cas des tournois non valués

Chaque arc permet de supprimer un certain nombre de 3-circuits. Appelons *degré* de l'arc  $(x, y)$  le nombre de circuits dans lequel il est impliqué ; c'est la terminologie consacrée dans l'hypergraphe  $H$  dont les sommets sont les arcs du tournoi et les arêtes les circuits considérés. Il est clair que l'on a intérêt à retourner les arcs de degré maximum puisque, s'ils appartiennent à des circuits disjoints, ils supprimeront autant de circuits que la somme de leurs degrés. Sans vérifier qu'ils soient disjoints, pour supprimer tous les  $NbC$  3-circuits de la partie  $F$ , il faudra inverser des arcs dont la somme des degrés est supérieure ou égale à  $NbC$ . On range donc les arcs dans l'ordre des degrés décroissants, et l'on somme ces degrés jusqu'à atteindre ou dépasser  $NbC$ . Le nombre d'arcs nécessaires constitue une nouvelle borne inférieure  $\chi(F)$  de la distance du sous-tournoi réduit à  $F$  à un ordre total.

Exemple : Sur le tournoi précédent, il y a 10 circuits de longueur 3,  $\{(a, b, d), (a, c, d), (a, f, d), (a, g, d), (b, c, e), (b, d, e), (b, g, e), (c, e, f), (c, g, f), (d, e, f)\}$ . L'arc  $(d, a)$  est de degré 4, l'arc  $(e, b)$  de degré 3 ; il y a 6 arcs de degré 2, et 11 arcs de degré 1. Pour supprimer les 10 circuits, il faudra donc au moins 4 arcs, si bien que tout ordre médian est au moins à distance 4.

En utilisant comme borne inférieure de la distance

$$V'''(O) = V(O) + \chi(F),$$

on développe une arborescence qui a 20% de sommets en moins qu'avec le paramètre  $\sigma$ . Cette économie se paye néanmoins par un temps de calcul plus long. Pour tout candidat au prolongement d'une section commençante, il faut dénombrer les 3-circuits de la partie complémentaire, en parcourant une liste pré-établie sur le tournoi complet, et calculer les degrés de leurs arcs, puis compter le nombre d'arcs des différents degrés.

#### 4. ET SI LE TOURNOI RESTE INCALCULABLE ?

Il reste dans ce cas une possibilité : restreindre le nombre de candidats et appliquer les mêmes procédures à un sous-tournoi. Bien évidemment, les gagnants de ce sous-tournoi ne sont pas nécessairement les premiers d'un ordre à distance minimum du tournoi complet (cela résulte de l'absence de la *propriété du sur-ensemble fort* ; voir par exemple [Laslier 1996] dans ce même numéro) ; il peut y avoir des arcs retour dont l'origine vient des candidats qui ont été éliminés. Néanmoins cette sélection peut se faire de façon raisonnable :

- en supprimant les candidats qui ne peuvent jamais être bien classés,
- en ne gardant que les premiers des chaînes transitives maximales (pour l'inclusion), c'est-à-dire les *vainqueurs de Banks* (voir [Banks 1985] ou [Laslier 1996]).

##### 4.1. Intervalles de rang dans les ordres à distance bornée

Si l'on connaît la distance  $\Delta$  d'un ordre solution, ou une approximation de cette distance, on peut déterminer un rang minimum (resp. maximum) pour chaque candidat. Ce rang est la plus petite (resp. grande) position qu'il peut occuper dans un ordre à distance inférieure ou égale à  $\Delta$ . Il n'est pas certain qu'il existe un ordre médian dans lequel ce candidat occupe cette position, mais on est assuré que dans tout ordre médian, il ne pourra pas être placé avant (resp. après) ce rang minimum (resp. maximum).

Considérons le candidat  $x_k$  d'un tournoi codé dans une matrice  $T$  comme précédemment. S'il peut être classé premier dans un ordre médian, c'est qu'on a  $\sum_{i \neq k} T(i, k) \leq \Delta$ . S'il peut être classé

second, après  $x_j$ , c'est qu'on a de même  $\sum_{i \in \{k, j\}} T(i, k) \leq \Delta$ . Or, il se peut que  $x_j$  réalise le

maximum des valeurs  $T(i, k)$ , (c'est-à-dire  $T(j, k) = \text{Max}_{i \neq k} T(i, k)$ ). Donc pour que  $x_k$  puisse être classé second, il faut avoir :

$$\sum_{i \neq k} T(i, k) - \text{Max}_{i \neq k} T(i, k) \leq \Delta.$$

De façon plus générale :

**PROPOSITION.** Soit  $x_k$  un candidat de  $T$  et soit  $I_p$  l'ensemble des indices des  $p$  plus grandes valeurs de  $T(i, k)$  ( $i \neq k$ ). Pour que le candidat  $x_k$  puisse être placé  $(p + 1)$ -ième dans un ordre médian, il faut avoir

$$\sum_{i \neq k} T(i, k) - \sum_{i \neq k \text{ et } i \notin I_p} T(i, k) \leq \Delta,$$

c'est-à-dire que la somme des  $n - (p + 1)$  plus petites valeurs  $T(i, k)$  (en ne considérant pas  $T(k, k)$ ) ne doit pas excéder  $\Delta$ .

En raisonnant de la même façon sur les lignes de la matrice  $T$ , si  $x_k$  est placé au rang  $p + 1$  dans un ordre à distance  $\Delta$ , c'est que la somme des  $p$  plus petites valeurs de  $\{T(k, i), i \neq k\}$  est inférieure ou égale à  $\Delta$ . Sinon on dépasse  $\Delta$  avec les seuls arcs retour d'origine  $x_k$ .

On peut donc calculer pour chaque candidat un rang minimum et un rang maximum dans un ordre médian. L'exemple suivant illustre ces considérations.

Exemple : Soit  $T$  le tournoi défini par la matrice ci-dessous :

$T$	$a$	$b$	$c$	$d$	$e$	$f$
$a$	-	2	0	0	3	0
$b$	0	-	5	2	4	1
$c$	2	0	-	0	7	0
$d$	5	0	6	-	2	0
$e$	0	0	0	0	-	2
$f$	3	0	9	4	0	-

L'ordre  $b > f > d > c > a > e$  est à distance 4 de  $T$ , d'où  $\Delta = 4$ . Dans un ordre à distance 4, il faut que  $d$  et  $f$  (ou  $d$  et  $e$ ) soient placés avant  $a$  ; il ne peut être au mieux que troisième. Il ne peut être dernier, car il serait à l'origine d'arcs retour dont la somme des poids est 5. Il est donc au pire cinquième. L'application de la proposition précédente donne les rangs minimum et maximum suivants :

	$a$	$b$	$c$	$d$	$e$	$f$
min	3	1	4	2	4	1
max	5	4	5	4	6	4

Pour le rang minimum, qui est le seul qui nous importe vraiment ici, il suffit de ranger pour tout  $k$  ( $1 \leq k \leq n$ ) les valeurs de  $\{T(i, k)\}_{i \neq k}$  dans l'ordre croissant et de sommer les valeurs dans cet ordre tant qu'on ne dépasse pas  $\Delta$ . Si l'on peut additionner  $p$  valeurs, il y a au plus  $n - p - 1$  candidats que l'on peut placer avant  $x_k$  pour qu'il puisse sortir ; il est donc au mieux au rang  $n - p$ .

Pour éliminer des candidats, il suffit de considérer la borne  $\Delta$  donnée par une méthode heuristique, et de retirer tous les candidats qui dépassent un rang seuil, par exemple ceux qui ne sont jamais dans la première moitié. Dans l'exemple précédent, on élimine  $c$  et  $e$ . Cette technique n'est vraiment efficace que sur les tournois valués, car avec les tournois non pondérés, on ne comptabilise que des valeurs unitaires. Ainsi avec la borne de 4 déterminée sur l'exemple de la partie 2, tous les candidats peuvent être premiers ! On n'aurait donc éliminé personne.

#### 4.2. Les premiers des chaînes transitives maximales

Une *chaîne transitive* est une partie de  $X$  totalement ordonnée qui ne présente aucun arc retour. Elle est *maximale* si elle n'est incluse dans aucune autre chaîne. Un sommet qui arrive en tête d'une chaîne transitive maximale s'appelle un *vainqueur de Banks* [1985]. Bien que l'ensemble des vainqueurs de Kemeny et celui des vainqueurs de Banks puissent être disjoints, même pour des tournois valués par 1 (voir par exemple [Charon, Hudry et Woïrgard 1996]), on peut calculer les premiers sur le sous-tournoi engendré par les seconds. En effet, une chaîne transitive maximale est un sous-ensemble de candidats au classement « indiscutable », chacun étant meilleur que tous ceux qui le suivent. La restriction du tournoi à leurs vainqueurs se justifie par l'observation que si  $x$  n'est jamais premier d'une telle chaîne, c'est qu'il existe toujours un élément  $y$  qui domine tous ceux que  $x$  précède dans un classement indiscutable.

Soulignons qu'un élément  $x$  peut ne pas être premier dans la chaîne  $(y, x, z, u)$  mais premier dans la chaîne  $(x, t, u)$  ; il suffit que  $y$  ne domine pas  $t$ . Puisque  $x$  est premier au moins une fois, il ne sera pas éliminé.

Pour déterminer les vainqueurs de Banks, on associe à chaque candidat  $x$  l'ensemble  $Nd$  des autres candidats qui ne le dominent pas dans  $T$ , c'est-à-dire l'ensemble de ses successeurs dans  $T$  y compris  $x$  lui-même :

$$Nd(x) = \{y \mid T(y, x) = 0\} \cup \{x\}.$$

Pour énumérer toutes les chaînes transitives maximales, on développe encore une arborescence. On place au premier niveau tous les éléments  $x$  tels que  $Nd(x)$  est une partie maximale (pour l'inclusion). À n'importe quel nœud correspond une chaîne  $P$  des éléments placés, que l'on obtient en remontant jusqu'à la racine. Pour prolonger cette chaîne :

- on construit la partie  $Q$  des éléments non placés qui ne dominent aucun des éléments de  $P$  :

$$Q = \bigcap_{x \in P} Nd(x) \setminus P$$

- on construit l'ensemble  $D$  des éléments qui dominent au moins un des éléments de  $P$  ou qui sont dans  $P$  :  $D = X \setminus Q$
- on prolonge  $P$  avec chaque élément  $y$  de  $Q$  tel que  $Nd(y) \setminus D$  est maximale ; autrement dit, on prolonge  $P$  par toutes les chaînes transitives maximales du sous-tournoi engendré par les sommets de  $Q$ .

On obtient ainsi toutes les chaînes transitives. Cet algorithme est plus compliqué que celui qui consiste à placer tous les éléments de  $Q$  après chaque nœud, mais on obtient dans ce cas un très grand nombre d'inclusions et une arborescence beaucoup plus grande. Même en ne prenant que



les éléments de  $Q$  qui correspondent à des parties maximales, il peut y avoir des inclusions et il faut comparer deux à deux les chaînes transitives obtenues pour ne conserver que les maximales.

**Exemple** : Sur le tournoi non valué du paragraphe 2, on obtient :  $Nd(a) = \{a, b, c, e, f, g\}$ ,  $Nd(b) = \{b, c, d, f, g\}$ ,  $Nd(c) = \{c, d, e, g\}$ ,  $Nd(d) = \{a, d, e\}$ ,  $Nd(e) = \{b, e, f\}$ ,  $Nd(f) = \{c, d, f\}$ ,  $Nd(g) = \{d, e, f, g\}$ . On a  $Nd(e) \subset Nd(a)$  et  $Nd(f) \subset Nd(b)$ , si bien que seuls  $a, b, c, d$  et  $g$  sont au premier niveau. Pour prolonger la chaîne  $P = (a)$ , on a  $Q = \{b, c, e, f, g\}$  et  $D = \{a, d\}$ , si bien qu'il y a 4 successeurs possibles pour prolonger  $a$  : ce sont  $b, c, e$  et  $g$  ( $f$  est éliminé par  $b$ ).

Ultérieurement, pour prolonger la chaîne  $P = (a, b)$  on a  $Q = \{c, f, g\}$  et  $D = \{a, b, d, e\}$ , il reste trois parties indépendantes, et donc 3 successeurs.

Au total on obtient 14 chaînes transitives, dont  $g > d > e$ ,  $g > e > f$  et  $g > f > d$  qui ne sont pas maximales, puisqu'elles sont respectivement incluses dans  $c > g > d > e$ ,  $a > g > e > f$  et  $b > g > f > d$ . Il en reste 11, toutes de longueur 4 sauf une :  $d > a > e$  ; les autres commencent par  $a$  :  $a > b > c > g$ ,  $a > b > f > c$ ,  $a > b > g > f$ ,  $a > c > g > e$ ,  $a > e > b > f$ ,  $a > g > e > f$ ,  $b$  :  $b > c > g > d$ ,  $b > f > c > d$ ,  $b > g > f > d$ , ou  $c$  :  $c > g > d > e$ . Donc si l'on devait sélectionner les candidats en tête de ces chaînes, on considérerait le sous-tournoi correspondant aux candidats  $\{a, b, c, d\}$ .

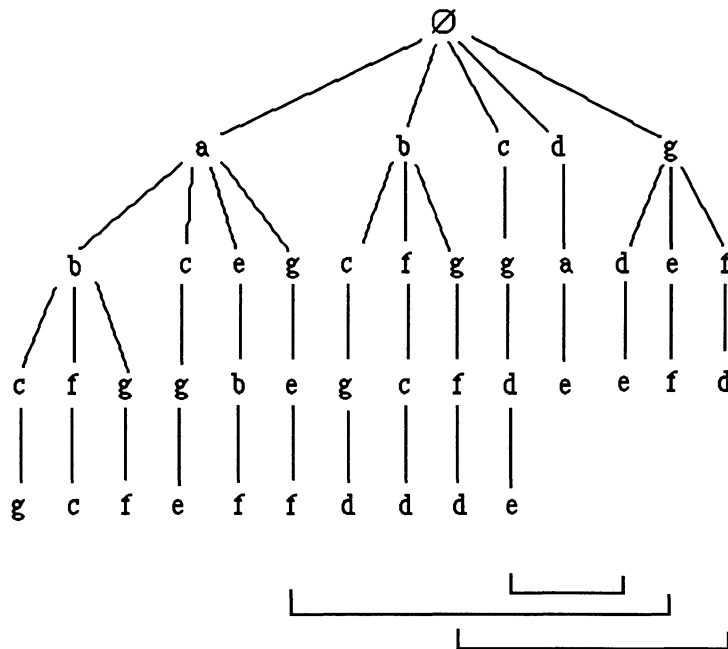


Figure 1 : Arbre des chaînes transitives ; celles commençant par  $g$  ne sont pas maximales

Cette technique de sélection préalable des vainqueurs de Banks est plus fondée pour les tournois non valués, car elle est indépendante des valuations. Si par exemple on donne la valeur 6 à tous les arcs situés au-dessous de la diagonale dans la table du tournoi de la partie 2, tous les ordres médians commencent par  $g$ , mais  $g$  n'est toujours pas le premier d'une chaîne transitive maximale.

## BIBLIOGRAPHIE

- BANKS, J. (1985) "Sophisticated voting outcomes and agenda control", *Social Choice and Welfare*, 2, 295-306.  
 BARTHÉLEMY, J.-P. et B. MONJARDET (1981) "The median procedure in cluster analysis and social choice theory", *Mathematical Social Sciences*, 1, 235-267.

- BARTHÉLEMY, J.-P., A. GUÉNOCHE et O. HUDRY (1989) "Median linear orders: heuristics and a branch and bound algorithm", *European Journal of Operational Research*, 41, 313-325.
- CHARON-FOURNIER, I., A. GERMA et O. HUDRY (1992) "Utilisation des scores dans des méthodes exactes déterminant les ordres médians de tournois", *Mathématiques, Informatique et Sciences Humaines*, 119, 53-74.
- CHARON, I., O. HUDRY et F. WOIRGARD (1996) "Ordres médians et ordres de Slater des tournois", *Mathématiques, Informatique et Sciences Humaines*, ce même numéro.
- CONDORCET, M.J.A.N. Caritat (marquis de) (1785) *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, Paris.
- GUÉNOCHE, A. (1977) "Un algorithme pour pallier l'effet Condorcet", *RAIRO*, 11, 1, 73-83.
- HUDRY, O. (1989) *Recherche d'ordres médians : complexité, algorithmique et problèmes combinatoires*, Thèse ENST, Paris.
- HUDRY, O. et F. WOIRGARD (1994) "Combinatorics and voting theory: on the number of median orders of tournaments", communication à "Conference on combinatorics in the behavioral sciences", 15-19 août 1994, Irvine, États-Unis.
- KEMENY, J.G. (1959) "Mathematics without numbers", *Daedalus*, 8, 577-591.
- LASLIER, J.-F. (1996) "Solutions de tournois : un spicilège", *Mathématiques, Informatique et Sciences Humaines*, ce même numéro.
- MONJARDET, B. (1973) "Tournois et ordres médians pour une opinion", *Mathématiques et Sciences Humaines*, 43, 55-70.
- REMAGE Jr., R. et W.A. THOMPSON Jr. (1966) "Maximum likelihood paired comparison rankings", *Biometrika*, 53, 143-149.
- SMITH, A.F.M. et C.D. PAYNE (1974) "An algorithm for determining Slater's  $i$  and all nearest adjoining orders", *British Journal of Mathematical and Statistical Psychology*, 27, 49-52.