Analyse harmonique et difficulté d'approximation

P. Pansu, Université Paris-Sud 11 et Ecole Normale Supérieure

11 mai 2011

Au menu aujourd'hui :

- Un théorème de la théorie du choix social.
- Sa preuve, qui utilise de l'analyse harmonique (sur l'hypercube, sur l'espace gaussien).
- Une application à la difficulté d'approximation, en informatique théorique.

Une *fonction booléenne* est une fonction $\{-1,1\}^n \to \{-1,1\}$. On peut y penser comme à un procédé pour agréger des votes, i.e. produire une décision à partir des votes de n électeurs.

Exemple

Le i-ème dictateur *est* $Dict_i(x) = x_i$. *La* majorité *est* $Maj(x) = signe(\sum x_i)$.

Un procédé d'agrégation devrait avoir les propriétés suivantes.

1. Aucun électeur ne joue de rôle prépondérant.
2. Des erreurs dans le dépouillement des votes ont peu de chance de faire basculer le résultat.

Définition

*L'*influence $Inf_i(f)$ *du i-ème électeur sur f est la probabilité que, lorsque le i-ème électeur change d'avis, la valeur de f change.*

$$Inf_i(f) = \mathbb{P}(f(xe_i) \neq f(x)),$$

où $e_i \in \{-1,1\}^n$ est le vecteur dont les coordonnées valent 1 sauf la i-ème.

Définition

La sensibilité au c-bruit *de f est la probabilité que, lorsque chaque vote est modifié indépendamment avec probabilité c, la valeur de f change.*

$$Sens_c(f) = \mathbb{P}_{x,z}(f(xz) \neq f(x)),$$

où les coordonnées $z_i \in \{-1, 1\}$ sont i.i.d., indépendantes de x, et $\mathbb{P}(z_i = -1) = c$.

Exemple

Le dictateur $Dict_i$ a $Inf_i(Dict_i) = 1$, $Inf_j(Dict_i) = 0$ si $j \neq i$, et

$$Sens_c(Dict_i) = c.$$

La majorité a $Inf_i(Maj) = \frac{2}{\sqrt{\pi n}}$ et

$$\lim_{n \to \infty} Sens_c(Maj) = \frac{1}{\pi} \arccos(1 - 2c).$$

Preuve : Théorème Central Limite.

De tous les procédés d'agrégation, la majorité est celui qui satisfait le mieux aux deux critères d'influence et de sensibilité ci-dessus. C'est la substance du Théorème MIS (`Majority is stablest`).

Théorème (Mossel, O'Donnell, Oleskiewicz 2005)

Soit $c \in [\frac{1}{2}, 1]$. De toutes les fonctions booléennes $\{-1,1\}^n \to \{-1,1\}$, dont les influences sont petites, Maj est celle dont la sensibilité au c-bruit est asymptotiquement la plus faible, lorsque n tend vers l'infini.

En fait, l'énoncé est non asymptotique : pour tout $\epsilon > 0$, il existe $\tau(\epsilon)$ tel que si toutes les influences $Inf_i(f) < \tau$, alors $Sens_c(f) \geq \frac{1}{\pi} \arccos(1-2c) - \epsilon$.

Dans la suite, on va

1. Donner un aperçu de la preuve.
2. Expliquer une conséquence : difficulté d'approximation pour le problème MAX CUT.

Preuve du Théorème MIS

1. Principe d'invariance : on remplace le cube $\{-1,1\}^n$ muni de la mesure de probabilité uniforme, par l'espace gaussien, i.e. l'espace euclidien $\mathbb{R}^n$ muni de la mesure gaussienne γ_n, de densité $(2\pi)^{-n/2}\exp(-|x|^2/2)$.
2. Dans l'espace gaussien, un argument de symétrisation dû à Ehrhard et Borell montre que parmi les fonctions de moyenne nulle, à valeurs dans $[-1,1]$, les fonctions "signe de forme linéaire" maximisent $Sens_c$, $c < \frac{1}{2}$.
3. Soit $c > \frac{1}{2}$. Si $g(x) = \frac{1}{2}(f(x) - f(-x))$, alors
$Sens_c(f) \geq Sens_c(g) = 1 - Sens_{1-c}(g) \geq 1 - \frac{1}{\pi}\arccos(1-c) - \epsilon = \frac{1}{\pi}\arccos(c) - \epsilon$.

Si $f : \{-1,1\}^n \to \mathbb{R}$, alors f s'étend à $\mathbb{R}^n$ de façon unique en un polynôme de degré partiel 1 (décomposition de Fourier-Walsh). Le principe d'invariance suivant généralise au cas des fonctions non linéaires le Théorème Central Limite. Il exprime quantitativement le fait que pour certaines fonctions f, les lois de f sur $\{-1,1\}^n$ et sur $\mathbb{R}^n$ gaussien sont voisines.

Proposition

Soit x tiré uniformément dans $\{-1,1\}^n$ et Y tiré indépendamment dans l'espace gaussien. Soit $f : \{-1,1\}^n \to \mathbb{R}$ de degré $\leq d$. On suppose que les influences satisfont $Inf_i(f) \leq \tau$. Alors pour toute fonction $\Psi : \mathbb{R} \to \mathbb{R}$ de classe C^4,

$$|\mathbb{E}(\Psi(f(x))) - \mathbb{E}(\Psi(f(Y)))| \leq \tau 10^d \parallel \Psi \parallel_{C^4} .$$

Preuve.

- Changer les variables une par une (d'où influence).
- Formule de Taylor, $\mathbb{E}(x_i^p) = \mathbb{E}(Y_i^p)$ pour $p \leq 3$.
- Inégalité d'hypercontractivité de Bonami pour estimer le reste.

Définition (A. Ehrhard 1983)

Soit $u \in L^2(\mathbb{R}^n, \gamma_n)$ une fonction positive. Sa symétrisée *est la fonction décroissante u^* définie sur $\mathbb{R}$ par*

$$\gamma_n(\{u > t\}) = \gamma_1(\{u^* > t\}).$$

Théorème (C. Borell 1985)

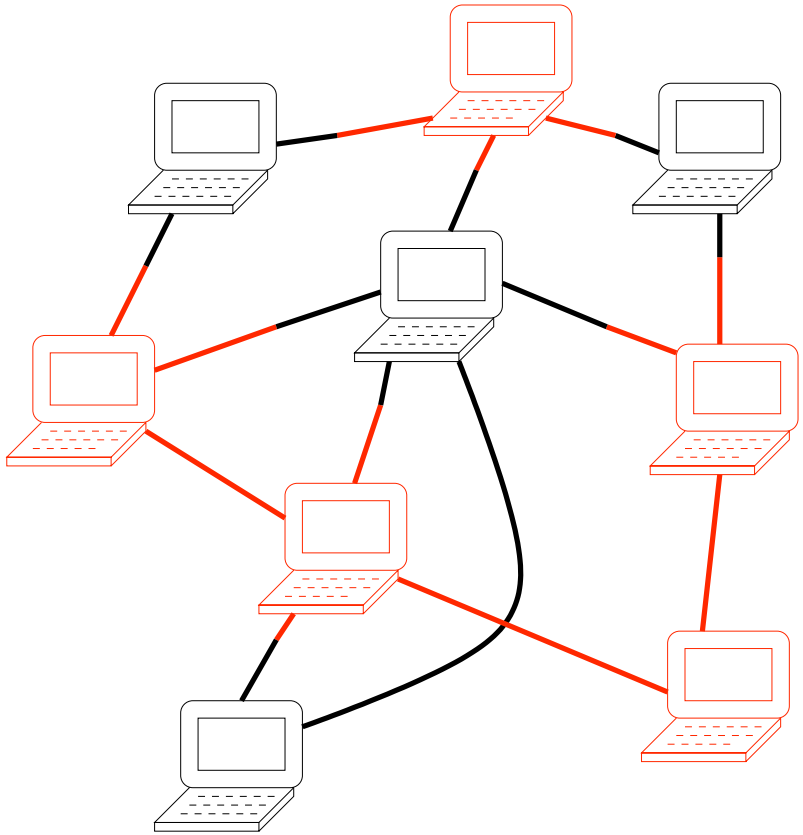
Soit $u \in L^2(\mathbb{R}^n, \gamma_n)$ une fonction positive. Alors pour $c < \frac{1}{2}$,

$$Sens_c(u) \geq Sens_c(u^*).$$

Preuve.

- Principe du maximum pour le semi-groupe d'Ornstein-Uhlenbeck U_t.
- $1 - 2Sens_c(u) = \langle U_t u, u \rangle$ pour $e^{-t} = 1 - 2c$.
- Inégalité isopérimétrique gaussienne.

Un virus se balade sur le réseau local du LAMFA. Déjà 5 machines infectées. On coupe les connexions qui les relient aux ordinateurs encore sains. Il y en a 9.

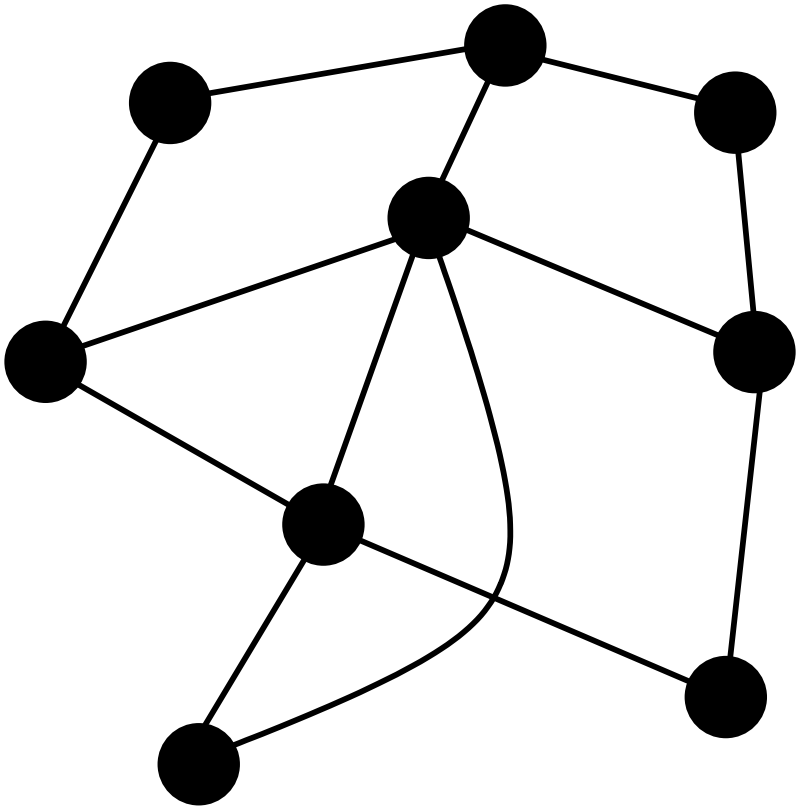


Un réseau n'est rien d'autre qu'un graphe : ici, 9 sommets, 13 arêtes. L'infection consiste à colorier une partie des sommets. La thérapie, à couper les arêtes bicolores. Ici, 9 arêtes coupées.

Il a suffi de couper 9 arêtes. Est ce que cela aurait pu être pire ? Existe t'il des infections qui nécessitent une thérapie plus radicale ?

Question

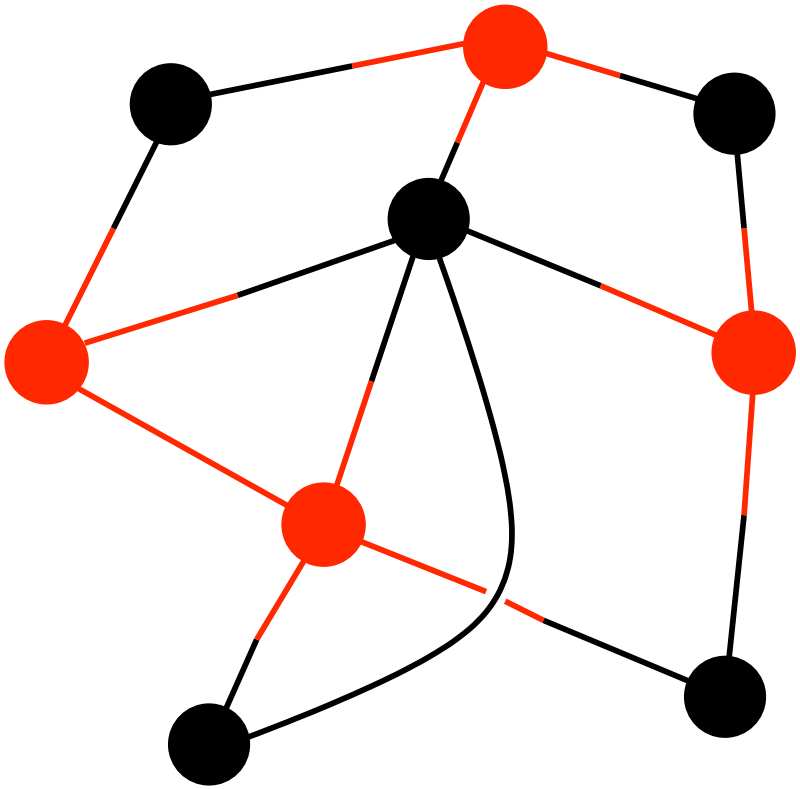
Existe-t'il un coloriage donnant plus de 9 arêtes bicolores ?



Il a suffi de couper 9 arêtes. Est ce que cela aurait pu être pire ? Existe t'il des infections qui nécessitent une thérapie plus radicale ?

Question

Existe-t'il un coloriage donnant plus de 9 arêtes bicolores ?



Oui, 11 arêtes. Mais jamais davantage.

Définition

Soit G un graphe pondéré (i.e. une distribution de probabilité sur les arêtes du graphe complet). On appelle coupe maximale(G) le maximum, sur tous les coloriages de G, de la probabilité qu'une arête soit coupée.

Le graphe étudié jusqu'à présent, muni de la pondération uniforme, a une coupe maximale de $\frac{11}{13}$.

Problème

MAX CUT : Ecrire un algorithme qui, étant donnés un graphe G à n sommets, et un nombre entier k, décide si la coupe maximale de G est $\geq k$ ou non.

Définition

Un problème est dit NP-complet *s'il est NP, et si tout problème NP s'y ramène en temps polynômial.*

Théorème (Cook, Levine 1971, Karp 1972)

MAX CUT est NP-complet.

Puisqu'on ne peut problement pas résoudre exactement le problème MAX CUT, on tente de le résoudre de façon approchée.

Définition

Soit $\alpha < 1$. Une résolution α-approchée *de MAX CUT, c'est un algorithme qui, étant donné un graphe pondéré G à n sommets, trouve en un temps polynômial en n, un coloriage dont le nombre d'arêtes bicolores est $\geq \alpha \times$ coupe maximale(G). Cet algorithme a le droit d'effectuer des tirages au hasard. On demande dans ce cas que le score annoncé soit obtenu avec probabilité $\geq \frac{1}{2}$.*

Exemple

Coloriage aléatoire. On tire indépendamment au hasard la couleur de chaque sommet.

Analyse : Chaque arête a une chance sur deux d'être bicolore. L'espérance du nombre d'arêtes bicolores vaut $\frac{1}{2} \times$ nombre d'arêtes $\geq \frac{1}{2} \times$ coupe maximale(G). Par symétrie, avec probabilité $\geq \frac{1}{2}$,

$$\text{nombre d'arêtes bicolores} \geq \frac{1}{2} \text{coupe maximale}(G).$$

Donc il s'agit d'une résolution $\frac{1}{2}$-approchée de MAX CUT.

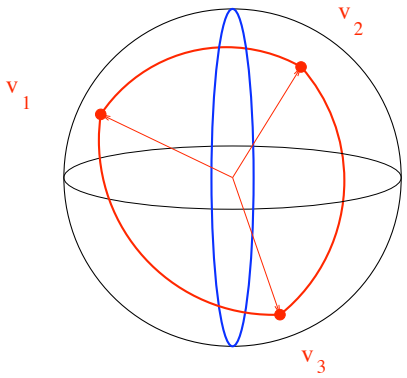
Théorème (Goemans-Williamson 1995)

Le problème MAX CUT possède une résolution 0.878...*-approchée.*

On dessine le graphe sur une sphère. Les sommets deviennent des vecteurs $v_1, \ldots, v_n$.
On les dispose de sorte à rendre la quantité suivante la plus grande possible

$$\mathrm{SDP} = \mathbb{E}_{\text{aretes } ij}\left(\frac{1}{2}(1 - v_i \cdot v_j)\right).$$

Il s'agit d'un problème de *programmation semi-définie*. La disposition optimale peut être calculée, avec une précision arbitraire, en temps polynômial.
L'idée est qu'à tout coloriage correspond une disposition particulière des vecteurs (superposés sur deux points diamétralement opposés), pour laquelle SDP est le nombre d'arêtes bicolores. En particulier, max SDP $\geq$ coupe maximale.
Inversement, on associe un coloriage à la disposition optimale comme suit : on coupe le graphe en deux au moyen d'un plan tiré au hasard.



Analyse. La probabilité que l'arête $v_i v_j$ soit coupée (bicolore) vaut $\frac{1}{\pi}\arccos(v_i \cdot v_j)$. L'espérance du nombre d'arêtes bicolores est

$$E = \mathbb{E}_{ij}(\frac{1}{\pi}\arccos(v_i \cdot v_j)).$$

Comme pour tout $x \in [0,1[$,

$$\frac{\frac{1}{\pi}\arccos(x)}{\frac{1}{2}(1-x)} \geq 0.878...,$$

$E \geq 0.878... \max SDP \geq 0.878... \times$coupe maximale.

Remarque (U. Feige, G. Schechtman 2002)

Il existe des graphes pour lesquels la méthode donne une coupe arbitrairement proche de $0.878... \times$ *coupe maximale.*

Théorème (S. Khot, G. Kindler, E. Mossel, R. O'Donnell 2005)

Sous une hypothèse un peu plus forte que P≠NP, 0.878... est le seuil optimal d'approximabilité pour MAX CUT. Autrement dit, si $\alpha > 0.8785672057848516...$, il n'existe pas de résolution α-approchée de MAX CUT.

La preuve est l'aboutissement d'idées qui remontent à Gödel (vérification de preuves, vérification probabiliste de preuves) et repose sur le théorème MIS.

L'hypothèse un peu plus forte que P≠NP, c'est la Conjecture des Jeux Uniques (`Unique Games Conjecture`) de Subhash Khot.

Définition

On s'intéresse à des jeux coopératifs à deux joueurs. Des couples de questions (i, j) sont tirés suivant une distribution connue. Les joueurs doivent donner des réponses $a = S(i)$ et $b = S(j)$ sans savoir quelle question a été posée à l'autre. Dans un jeu projectif, *pour chaque question i, chaque question j et chaque réponse possible a du premier joueur, il y a une unique réponse $b = \pi_{ij}(a)$ du second qui les fait gagner tous les deux. Les joueurs cherchent une stratégie commune S qui maximise la probabilité de gain,*

$$Valeur\ du\ jeu = \max_S \mathbb{P}_{(i,j)}(S(j) = \pi_{ij}(S(i))).$$

Théorème

Soit n le nombre de questions et k le nombre de réponses par question. Pour tout $\epsilon > 0$, il existe k tel qu'il est NP-difficile de décider entre les deux cas de figure suivants (sachant qu'on est dans l'un des deux).

1. *La valeur du jeu est 1.*
2. *La valeur du jeu est $< \epsilon$.*

C'est une conséquence folklorique du théorème PCP (`Probabilistically Correct Proofs`) de Arora, Safra et al. (1992), et du théorème de répétition parallèle de Ran Raz (1995).

Définition

Un jeu unique *est un jeu projectif dans les deux sens, i.e. pour chaque question i, chaque question j et chaque réponse possible b du second joueur, il y a aussi une unique réponse $a = \pi_{ij}^{-1}(b)$ du premier qui les fait gagner.*

Il est facile de décider si la valeur du jeu vaut 1. En effet, dans ce cas, un choix quelconque de réponse à une question détermine uniquement les réponses à toutes les autres. Il suffit de vérifier la cohérence des réponses obtenues.En revanche, il semble difficile de décider si la valeur du jeu est proche de 1 ou non.

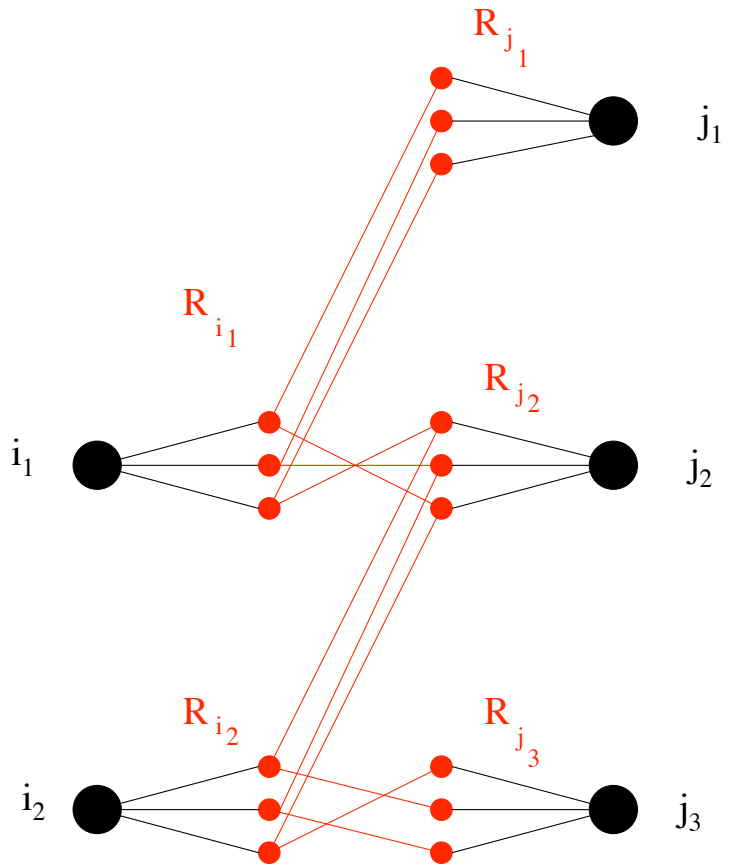
Conjecture (S. Khot 2002)

Pour tout $\epsilon > 0$ et tout $\delta > 0$, il existe k tel qu'il est NP-difficile de décider entre les deux cas de figure suivants (sachant qu'on est dans l'un des deux).

1. *La valeur du jeu est $> 1 - \delta$.*
2. *La valeur du jeu est $< \epsilon$.*

Les opinions sont partagées sur cette conjecture. En revanche, personne ne parie sur le fait que le problème des jeux uniques est dans P.

Un jeu unique J peut être vu comme un graphe biparti pondéré à n sommets, muni d'un ensemble R_i à k éléments attaché à chaque sommet i, et d'une bijection $\pi_{ij} : R_i \to R_j$ pour chaque arête ij. Une stratégie des deux joueurs se traduit par un système de réponses $j \mapsto S(j) \in R_j$. La valeur du jeu J est le maximum (parmi tous les choix de S) de la probabilité (sous la distribution donnée sur les arêtes) que $S(j) = \pi_{ij}(S(i))$.

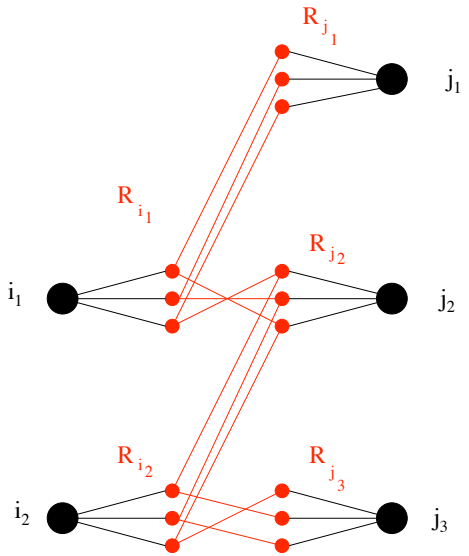


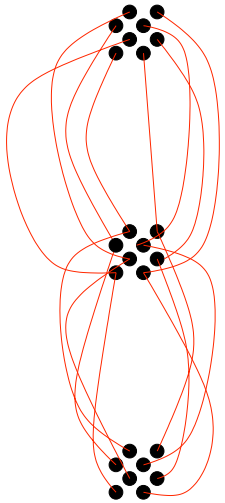
Un jeu unique J peut être vu comme un graphe biparti pondéré à n sommets, muni d'un ensemble R_i à k éléments attaché à chaque sommet i, et d'une bijection $\pi_{ij} : R_i \longrightarrow R_j$ pour chaque arête ij. Une stratégie des deux joueurs se traduit par un système de réponses $j \mapsto S(j) \in R_j$. La valeur du jeu J est le maximum (parmi tous les choix de S) de la probabilité (sous la distribution donnée sur les arêtes) que $S(j) = \pi_{ij}(S(i))$.

On associe à cette donnée J un graphe pondéré G, de sorte que si la valeur du jeu J est $> 1 - \delta$ (resp. $< \epsilon$), alors la coupe maximale de G est $> c - \delta'$ (resp. $< s + \epsilon'$). Un algorithme calculant en temps polynômial coupe maximale(G) à un facteur multiplicatif $\frac{s+\epsilon'}{c-\delta'}$ près décide si coupe maximale$(G) > c - \delta'$ ou si coupe maximale$(G) < s + \epsilon'$, donc décide si la valeur de J est $> 1 - \delta$ ou $< \epsilon$. Ce qui est interdit par la conjecture des jeux uniques. On conclut qu'il n'existe pas d'algorithme d'α-approximation de MAX CUT pour $\alpha > \frac{s}{c}$.

Pour tout $c > \frac{1}{2}$, la construction donne $s = \frac{1}{\pi} \arccos(1 - 2c)$. Le facteur optimal d'approximation est donc $\min_{c>1/2} \frac{1}{\pi c} \arccos(1 - 2c) = 0.878567205784851 6....$

L'ensemble des sommets du graphe pondéré G est la réunion des hypercubes $\{-1,1\}^{R_j}$. On définit une distribution de probabilité sur les arêtes xx', $x \in \{-1,1\}^{R_j}$, $x' \in \{-1,1\}^{R_{j'}}$, comme suit. On tire les arêtes ij et ij' au hasard suivant la distribution donnée, on tire x uniformément au hasard dans $\{-1,1\}^{R_j}$. On modifie $x' = \pi_{ij'} \circ \pi_{ij}^{-1}(x) \in \{-1,1\}^{R_{j'}}$ en changeant indépendamment chacun de ses bits avec probabilité c.





L'ensemble des sommets du graphe pondéré G est la réunion des hypercubes $\{-1,1\}^{R_j}$. On définit une distribution de probabilité sur les arêtes xx', $x \in \{-1,1\}^{R_j}$, $x' \in \{-1,1\}^{R_{j'}}$, comme suit. On tire les arêtes ij et ij' au hasard suivant la distribution donnée, on tire x uniformément au hasard dans $\{-1,1\}^{R_j}$. On modifie $x' = \pi_{ij'} \circ \pi_{ij}^{-1}(x) \in \{-1,1\}^{R_{j'}}$ en changeant indépendamment chacun de ses bits avec probabilité c.

A une stratégie S des deux joueurs correspond un coloriage des sommets de G : le sommet $x \in \{-1,1\}^{R_j}$ est colorié par sa $S(j)$-ème coordonnée (coloriage dictatorial). Si c est égal à 1 (i.e. tous les bits de y sont renversés), la probabilité que l'arête xx' soit coupée est égale à la probabilité que $S(j') = \pi_{ij'} \circ \pi_{ij}^{-1}(S(j))$, i.e. à la valeur de la stratégie S. Lorsque $c \neq 1$, cette probabilité est au pire multipliée par c, d'où coupe maximale$(G) \geq c$ valeur(J).

Inversement, un coloriage de G induit une famille $f_j : \{-1,1\}^{R_j} \to \{-1,1\}$ de fonctions booléennes. Si la probabilité qu'une arête soit coupée est $> s$, alors en moyenne, la sensibilité au c-bruit de f_j est $> s$. D'après le théorème MIS, si $s > \frac{1}{\pi} \arccos(1-2c) + \epsilon'$, au moins une des coordonnées a une forte influence sur f_j, d'où une stratégie S : tirer $S(j)$ au hasard parmi les coordonnées d'influence $> \tau$. On vérifie que la valeur de cette stratégie n'est pas très faible : valeur$(J) > \epsilon = \epsilon' \tau^2 / k$.