MATHÉMATIQUES ET SCIENCES HUMAINES

B. LECLERC

Les arbres et les indices de centralité et de compacité de P. Parlebas

Mathématiques et sciences humaines, tome 39 (1972), p. 27-35
<http://www.numdam.org/item?id=MSH_1972__39__27_0>

Math. Sci. hum. (10e année, n° 39, 1972, p. 27-35)

LES ARBRES ET LES INDICES DE CENTRALITÉ ET DE COMPACITÉ DE P. PARLEBAS

par

B. LECLERC [1]

Dans cet article, nous donnons la démonstration d'une conjecture de Parlebas sur la nature du graphe le plus centralisé et nous abordons l'étude de quelques aspects des indices qu'il propose (*cf.* « Centralité et compacité d'un graphe», dans ce numéro). Après avoir donné quelques définitions au paragraphe 1, nous nous intéressons, au paragraphe 2, à la compacité. Après avoir montré le rôle important joué par les arbres, nous donnons une démonstration, différente de celle de Parlebas, du fait que l'arbre-chaîne est le graphe connexe à n sommets le moins compact, l'intérêt de cette démonstration étant qu'elle donne le comportement de l'indice de compacité lorsque l'on fait subir à un arbre certaines transformations. Au paragraphe 3, nous montrons que les arbres jouent également un grand rôle du point de vue de la centralité, et nous étudions les sommets centraux des arbres. La notion de sommet central est naturelle, mais semble beaucoup moins connue que celle de centre d'un graphe. Dans le cas d'un arbre, elle généralise à un certain point de vue la médiane d'un ordre total. Enfin, au paragraphe 4, nous démontrons la conjecture de Parlebas.

Les numéros des références bibliographiques renvoient aux références données à la fin de l'article de Parlebas.

1. Définitions

Nous appelons *graphe* non orienté, un triplet $G = (X, U, \mathrm{i})$ où X et U sont deux ensembles finis, respectivement l'ensemble des *sommets* et l'ensemble des *arêtes* du graphe ; i est une application de U dans $P_2(X)$, l'ensemble des parties de X de cardinal 2. Soit u un élément de U. Les deux sommets de i (u) sont les *extrémités* de l'arête u. Avec ces définitions, il ne peut y avoir de boucle (arête dont les extrémités sont égales). On pose **card** $X = n$.

Un *graphe partiel* $G' = (X, U', \mathrm{i}')$ de G est obtenu par *suppression* d'arcs de G. L'ensemble des sommets de G' est X, l'ensemble de ses arêtes est $U' \subset U$, et l'application i' est la restriction de i à U'.

Une *chaîne* entre deux sommets x et y est un ensemble d'arêtes de G que l'on peut ordonner totalement de façon que toute arête autre que la première et la dernière, a une extrémité commune avec la précédente et l'*autre* commune avec la suivante. $\mathscr{C}_G(\mathrm{x}, \mathrm{y}) = \mathscr{C}_G(\mathrm{y}, \mathrm{x})$ est l'ensemble des chaînes entre x et y.

On appelle graphe *connexe* un graphe pour lequel entre deux sommets quelconques, il y a au moins une chaîne:

1. Centre de Mathématique Sociale, EPHE, 6e Section.

$$\forall\, x, y \in X, \qquad \mathscr{C}_G(x, y) \neq \emptyset.$$

Un *arbre* est un graphe connexe minimal, c'est-à-dire que G est un arbre si, et seulement si tout graphe partiel de G autre que G est non connexe. Il est classique que ceci équivaut à d'autres propriétés. Nous utiliserons l'une d'elles:

G est un arbre si et seulement si, entre deux sommets quelconques x et y, il y a une chaîne et une seule.

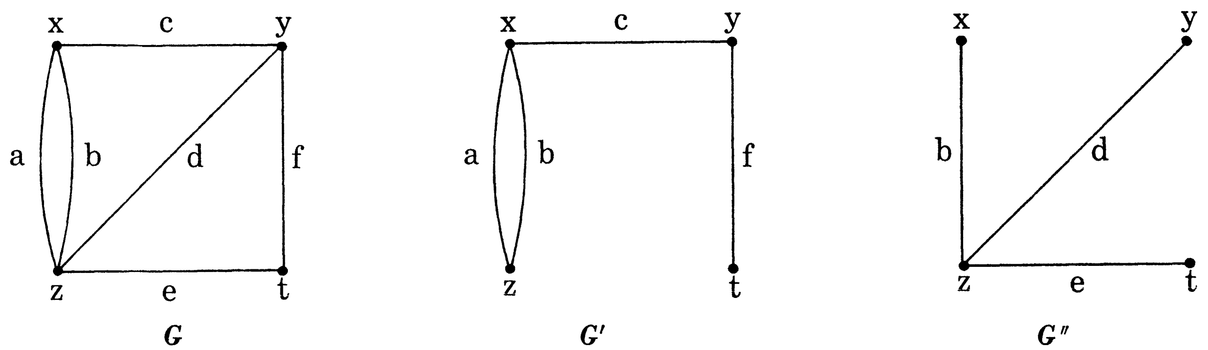


Figure 1

Un graphe G et deux graphes partiels G' et G'' de G. G'' est un arbre

Soit une application $l: U \to R^+ - \{O\}$. l (u) est la longueur de l'arête U. Si C est une chaîne quelconque de G, on définit l (C):

$$l(C) = \sum_{u \in C} l(u)$$

La *distance* d (x, y) sur X que nous employerons est la « distance du plus court chemin», définie par:

$$d(x, y) = \min_{C \in \mathscr{C}_G(x,y)} l(C)$$

avec pour tout $x \in X$, $d(x, x) = 0$.

d (x, y) est bien définie lorsque G est connexe.

Nous nous placerons souvent dans le cas où $l(u) = 1$, quelle que soit l'arête u. On a alors, si C est un chemin de G:

$$l(C) = \text{card } C$$

A tout sommet x on attache la quantité:

$$s(x) = \sum_{y \in X} d(x, y)$$

Un *sommet central* du graphe G est un sommet m tel que:

$$s(m) = \min_{x \in X} s(x)$$

On associe au graphe G la quantité:

$$s_G = \sum_{x \in X} s(x) = \sum_{x \in X} \sum_{y \in X} d(x, y),$$

si m est un sommet central, on a donc:

$$s_G \geqslant n\, s(m).$$

Nous nous intéressons au comportement de cette quantité s_G et de *l'indice de centralité* $C_G = \frac{s(m)}{s_G}$ dans la classe des graphes connexes à n sommets, lorsque la longueur de toute arête est égale à l'unité.

2. Comportement de s_G

Le résultat principal est celui-ci:

Proposition 1. Si G' est un graphe partiel connexe de G, on a $s_{G'} \geqslant s_G$.

C'est assez évident. On a en effet:

$$\mathscr{C}_{G'}(x, y) \subset \mathscr{C}_G(x, y)$$

d'où:

$$\min_{C \in \mathscr{C}_G(x, y)} l(C) \leqslant \min_{C \in \mathscr{C}_G(x, y)} l(C)$$

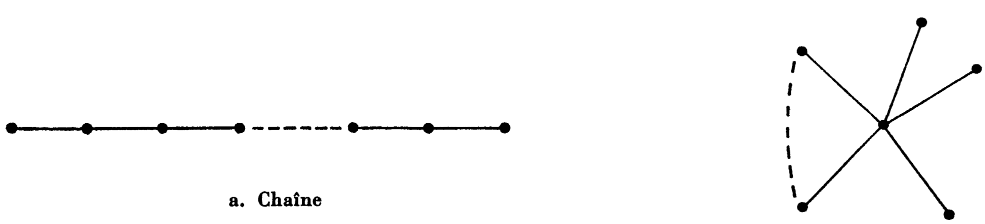
Donc, si d' est la distance du plus court chemin dans le graphe:

$$\forall\, x, y \in X \qquad d'(x, y) \leqslant d(x, y)$$

Intuitivement, en supprimant des routes, on ne peut qu'allonger les trajets. On note que si i est injective (il n'y a pas deux arêtes ayant les mêmes extrémités), on a $s_{G'} > s_G$, lorsque G' n'est pas égal à G, car si on enlève l'arête d'extrémités x et y, la distance entre x et y est augmentée.

Un résultat classique est que tout graphe connexe admet un graphe partiel connexe qui est un arbre. On déduit donc de la proposition 1 que les graphes connexes à n sommets maximisant s_G sont des arbres.

Montrons maintenant que ce sont ceux dont la configuration est « en chaîne » (Fig. 2a).



a. Chaîne

b. Étoile

Figure 2

Deux types d'arbres particuliers

Formellement, un arbre « en chaîne» est un arbre dont l'ensemble d'arêtes est lui-même une chaîne, un arbre « en étoile» est un arbre dont les chaînes sont de cardinal au plus égal à 2. On peut aussi les définir par leur nombre de sommets pendants (incidents à une seule arête) : deux pour l'arbre en chaîne, $n - 1$ pour l'étoile.

Pour montrer que s_G est maximal lorsque G est un arbre en chaîne, nous allons étudier un type de transformation que l'on peut faire subir à tout arbre G d'un autre type pour obtenir un nouvel arbre G' avec un sommet terminal de moins. En ce sens, on se rapproche donc de la chaîne.

Pour ne pas trop alourdir cette note, nous ne formaliserons pas complètement la définition de cette transformation (Fig. 3).

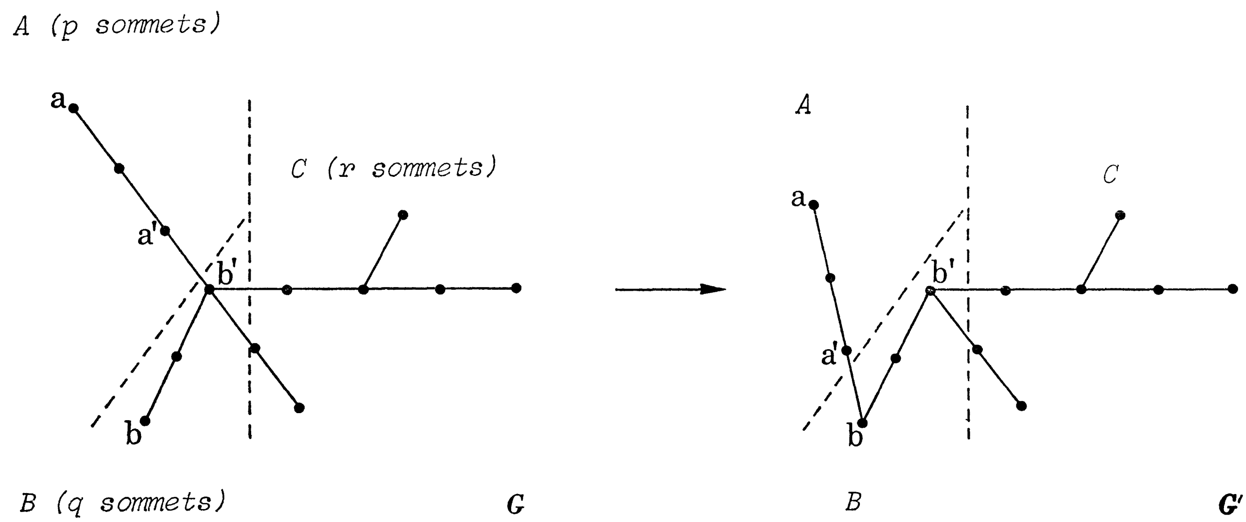


Figure 3

On partitionne X en trois classes A, B et C de cardinaux respectifs p, q et r. Les sommets de A sont tous incidents à deux arêtes sauf un sommet pendant a. Les sommets de B sont incidents à deux arêtes, sauf un sommet pendant b et un sommet b' adjacent à la fois à un sommet a' de A, un sommet de B et au moins un sommet de C (dans un arbre avec plus de deux sommets pendants, on peut toujours trouver de tels ensembles A, B et C). $C = X - A - B$, c'est-à-dire que $\{A, B, C\}$ est une partition de X.

La transformation (Fig. 3) consiste à remplacer l'arête { a', b' } par une arête { a', b }. Soient d et d' les distances du plus court chemin dans les arbres G et G'.

$$\text{si} \quad x \in C, y \in C, \qquad d'(x, y) = d(x, y)$$

$$\text{si} \quad x \in C, y \in B, \qquad d'(x, y) = d(x, y)$$

$$\text{si} \quad x \in C, y \in A, \qquad d'(x, y) = d(x, y) + q - 1$$

enfin:

$$\sum_{x \in A \cup B} \sum_{y \in A \cup B} d'(x, y) = \sum_{x \in A \cup B} \sum_{y \in A \cup B} d(x, y)$$

(il s'agit de la même configuration: chaîne à $p + q$ sommets). Par conséquent: $s'_G > s_G$.

Par suite de transformations analogues, on réduit jusqu'à deux le nombre de sommets pendants. On a alors une chaîne à n sommets qui maximise bien la quantité s_G.

Proposition 2. Parmi les graphes connexes non orientés à n sommets, l'arbre en chaîne est celui qui maximise s_G.

Nous venons donc de montrer que l'arbre en chaîne est le graphe à n sommets *le moins compact* au sens que vient de définir Parlebas.

3. Médiane d'un arbre

Proposition 3. Soit G un graphe connexe à n sommets, et m un sommet central de G. Il existe un graphe partiel G' de G vérifiant:

(i) G' est un arbre,

(ii) $\forall$ x $\in G$, d' (m, x) $= d$ (m, x).

Cette proposition est vraie dans le cas général où la longueur l (u) d'une arête n'est pas constante. m reste un sommet central de G' puisque on a: d' (x, y) $\geqslant d$ (x, y), $\forall$ x, y $\in X$.

Ceci est démontré dans [1], ch. 4, paragraphe 2, p. 57, dans le cas général où m est un sommet quelconque.

Considérons l'indice C_G de centralité du graphe G, tel qu'il a été défini précédemment par Parlebas.

$$C_G = \frac{s\,(\mathrm{m})}{s_G}.$$

Alors, pour l'indice de centralité $C_{G'}$, de G', on a, d'après la proposition 1:

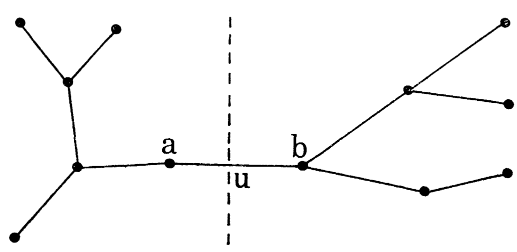
$$C_{G'} = \frac{s\,(\mathrm{m})}{s_{G'}} \leq \frac{s\,(\mathrm{m})}{s_G} = C_G.$$

On en déduit que les graphes connexes à n sommets « les plus centralisés » au sens de Parlebas, sont nécessairement des arbres. Nous montrons au paragraphe 4, que ce sont les arbres de configuration « en étoile ». Préliminairement, nous étudions ici la situation des sommets centraux dans les arbres, à partir du résultat suivant, vrai quelles que soient les longueurs des arcs (strictement positives).

Proposition 4. Soit G un arbre à n sommets, u une arête de G d'extrémités a et b, G_{a} et G_{b} les deux sous-arbres de G, à respectivement p et q sommets obtenus en supprimant l'arête u dans G (Fig. 4). a est un sommet de G_{a}, b un sommet de G_{b}.

Alors, dans l'arbre G, on a:

$$s\,(\mathrm{a}) - s\,(\mathrm{b}) = (q - p)\; l\,(\mathrm{u}).$$



G_{a} (p sommets) G_{b} (q sommets)

Figure 4

Si x est un des p sommets de G_a:

$$d\,(\mathrm{a}, \mathrm{x}) = d\,(\mathrm{b}, \mathrm{x}) - l\,(\mathrm{u})$$

et si x est un des q sommets de G_b:

$$d\,(\mathrm{a}, \mathrm{x}) = d\,(\mathrm{b}, \mathrm{x}) + l\,(\mathrm{u})$$

d'où la proposition 4, dont une conséquence importante est:

$$s\,(\mathrm{a}) \leqslant s\,(\mathrm{b}) \Leftrightarrow p \leqslant q$$

les inégalités devenant simultanément des égalités.

Cette démonstration est la même que celle qui permet de situer la ou les *médianes* d'un ordre total fini. Ainsi, dans le cas d'un arbre, la notion de point central est une généralisation de celle de médiane d'un ordre total. Nous parlerons maintenant de médiane d'un arbre. Donnons une caractérisation simple de ces médianes.

Proposition 5. Si et seulement si m est une médiane de l'arbre G, lors de la suppression d'une arête incidente à m, celui des deux sous-arbres obtenus qui contient m a un nombre de sommets au moins égal au nombre de sommets de l'autre.

Sinon, pour l'autre sommet x extrémité de u, on aurait s (x) $<$ s (m), et m ne serait pas médiane. Si les deux arbres ont même nombre de sommets, x est une seconde médiane de G. D'autre part, soit un sommet y non adjacent à m, et soit C la chaîne (unique) entre y et m, avec v l'arête de C incidente à y et u celle incidente à m (Fig. 5).

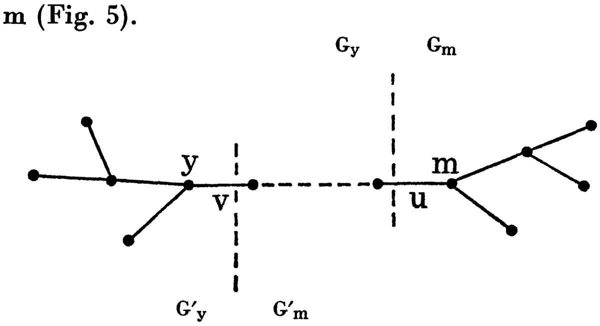


Figure 5

Puisque m est médiane, quand on supprime l'arête u dans G le sous-arbre G_y obtenu qui contient y a au plus autant de sommets que celui G_m, qui contient m. Si c'est l'arête v que l'on supprime dans G, on obtient deux sous-arbres G'_y et G'_m contenant respectivement y et m. Alors, comme il y a au moins un sommet intermédiaire entre m et y, G'_y contient strictement moins de sommets que G_y, donc que G_m, qui contient moins de sommets que G'_m et d'après la proposition 4, y ne peut être médiane.

Un arbre admet donc soit une médiane unique, soit deux médianes adjacentes, et ce dernier cas ne peut se présenter que si n est pair. On est dans une situation voisine, mais un peu différente de celle des ordres totaux finis pour lesquels il y a une médiane unique si n est impair, deux médianes voisines *si* et seulement si n est pair.

Nous avons cherché si la médiane d'un arbre généralisait en d'autres sens la médiane d'un ordre total. Ainsi, on peut se demander si elle se trouve sur le plus grand nombre de chaînes différentes de l'arbre. La Figure 6 donne un contre-exemple:

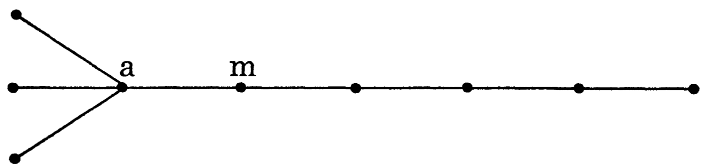


Figure 6

La médiane m est sur 24 chaînes différentes de l'arbre (compris celles dont elle est une extrémité) alors que le sommet a est sur 26 chaînes.

On peut aussi se demander si la médiane n'est pas un sommet situé sur le plus grand nombre de sous-arbres de l'arbre, (parties convexes de l'arbre), ceci nous ayant été suggéré par une remarque de Boulaye [9], p. 154. La Figure 7 nous donne encore un contre-exemple.

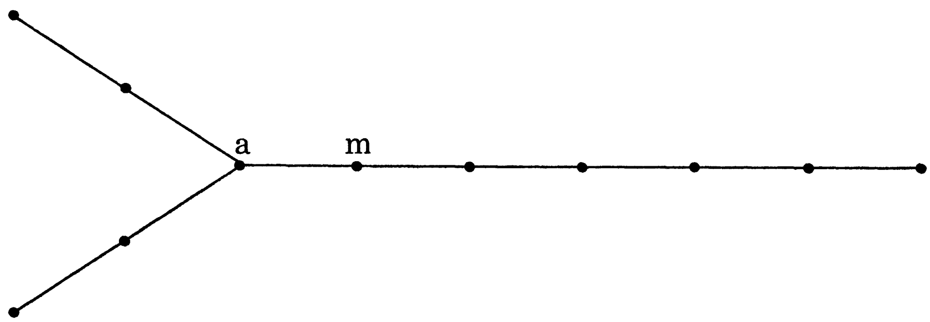


Figure 7

Le nombre de sous-arbres sur lesquels est situé le sommet a, mais pas la médiane m, est le nombre de chaînes passant par a de l'arbre contenant a obtenu en supprimant l'arête { a, m } (Fig. 8). Il est égal à 9 (compris l'arbre réduit au seul sommet a).

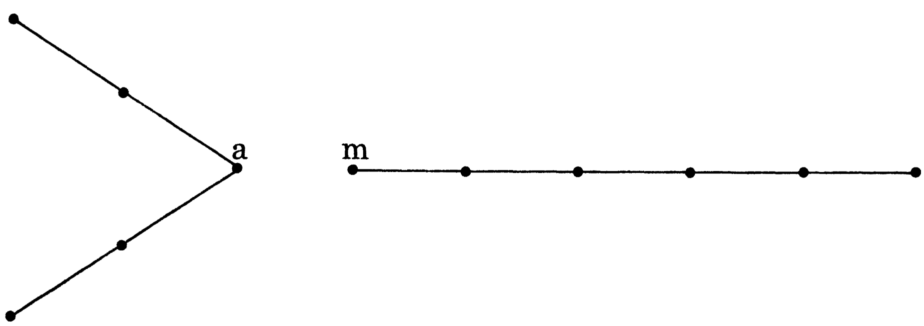


Figure 8

Par contre, le nombre de sous-arbres contenant m et non a, est égal à 6. Il y a donc plus de sous-arbres contenant a que de sous-arbres contenant m.

4. Les graphes les plus centralisés

Proposition 6. Lorsque $l(u) = 1$, quelle que soit l'arête u de G, les graphes connexes à n sommets qui minimisent l'indice C_G sont les arbres en étoile.

Rappelons d'abord (*cf. supra*, article de Parlebas) que l'indice de l'étoile à n sommets est égal à $\frac{1}{2(n-1)}$.

Comme pour la démonstration de la maximalité de S_G quand G est un arbre-chaîne, nous allons étudier le comportement de C_G lorsqu'on fait subir une transformation d'un certain type à un arbre G différent de l'étoile pour obtenir un nouvel arbre G'. Soit m une médiane de G; si G n'est pas une étoile, il existe au moins un sommet a à distance 2 de m. Soient b le sommet situé entre m et a, u l'arête d'extrémités a et b, v l'arête d'extrémités b et m (Fig. 9). La transformation étudiée consiste à remplacer l'arête u par l'arête u′ d'extrémités a et m.

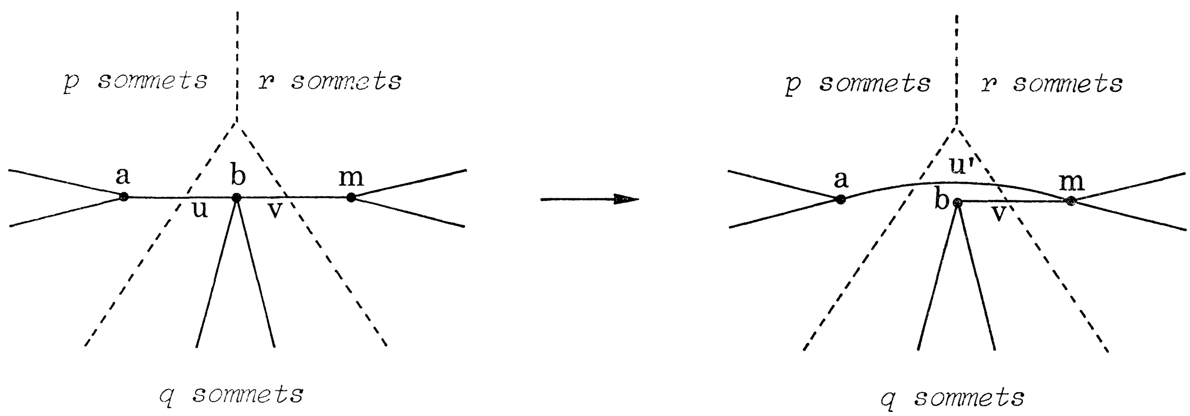


Figure 9

Si on supprime les arêtes u et v dans G (ou u′ et v dans G'), on obtient trois sous-arbres de G (ou de G') dont les ensembles de sommets sont respectivement A, B et C, avec $a \in A$, $b \in B$, $m \in C$.

On pose **card** $A = p$, **card** $B = q$, **card** $C = r$. p, q et r sont tous trois non nuls, $p + q + r = n$, car $\{A, B, C\}$ est une partition de X.

La proposition 5 permet de montrer que m est encore une médiane de G', et que si b est une médiane de G, ce n'en est plus une de G'.

Soit d' la distance du plus court chemin dans G'. $d'(x, y)$ est égal à $d(x, y)$ sauf dans les cas suivants:

$$\text{si } x \in A, y \in B \text{ ou si } x \in B, y \in A, \quad d'(x, y) = d(x, y) + 1$$

$$\text{si } x \in A, y \in C \text{ ou si } x \in C, y \in A, \quad d'(x, y) = d(x, y) - 1.$$

On en déduit que:

$$s'(m) = \sum_{x \in X} d'(m, x) = s(m) - p$$

$$s_{G'} = \sum_{x \in X} \sum_{y \in X} d'(x, y) = s_G - 2p(r - q).$$

Puisque $p \geqslant 1$, on a $s'(\mathrm{m}) > s(\mathrm{m})$. Une suite d'opérations analogues en choisissant s'il y a lieu une des deux médianes de l'arbre initial, aboutit donc finalement à obtenir l'arbre en étoile, le seul qui ait un sommet de centralité minimale $n - 1$. D'autre part, la proposition 5 entraîne que r est strictement supérieur à q, d'où $s_{G'} < s_G$. Comparons maintenant $C_{G'}$ et C_G.

$$C_G - C_{G'} = \frac{s(\mathrm{m})}{s_G} - \frac{s'(\mathrm{m})}{s_{G'}} = \frac{p\,(s_G - 2\,(r - q)\,s(\mathrm{m}))}{s_G\,(s_G - 2\,p\,(r - q))}$$

$C_G - C_{G'}$ est donc positif si et seulement si:

$$C_G = \frac{s(\mathrm{m})}{s_G} < \frac{1}{2\,(r - q)}.$$

Nous ne sommes pas parvenus à démontrer cette inégalité qui établirait que l'opération étudiée transforme toujours un arbre en un autre arbre plus centralisé. Nous n'avons pas non plus trouvé de contre-exemple. La proposition 6 est quand même vraie, car on a: $r \leqslant n - 2$ et $q \geqslant 1$, donc:

$$r - q \leqslant n - 3.$$

Donc si un arbre G a un indice C_G inférieur ou égal à celui de l'étoile, c'est-à-dire à $\frac{1}{2\,(n - 1)}$, on a:

$$C_G \leqslant \frac{1}{2\,(n - 1)} < \frac{1}{2\,(n - 3)} \leqslant \frac{1}{2\,(r - q)}.$$

On peut donc encore augmenter, par la transformation étudiée ci-dessus, la centralité de tout arbre autre que l'étoile, mais au moins aussi centralisé que l'étoile. Il y a là une contradiction, car en augmentant par étapes successives cette centralité, c'est-à-dire en diminuant C_G, on aboutit justement à l'étoile.

La conclusion est qu'il n'y a pas d'arbre au moins aussi centralisé (au sens de Parlebas) que l'étoile, donc pas de graphe connexe en général, d'après ce qui a été dit au début du paragraphe 3, ce qui établit la proposition 6.