MATHÉMATIQUES ET SCIENCES HUMAINES

E. JACQUET-LAGRÈZE

Analyse d'opinions valuées et graphes de préférence

Mathématiques et sciences humaines, tome 33 (1971), p. 33-55

<http://www.numdam.org/item?id=MSH_1971__33__33_0>

Math. Sci. hum. (9e année, n° 33, 1971, p. 33-55)

ANALYSE D'OPINIONS VALUÉES ET GRAPHES DE PRÉFÉRENCE [1]

par

E. JACQUET-LAGRÈZE [2]

INTRODUCTION

L'objet de cet article est de développer une analyse mathématique sur un ensemble d'opinions valuées sur des objets (ou des options) qui se trouvent comparés paire par paire.

Deux approches sont généralement envisagées. La première correspond au problème de l'agrégation des préférences. Il s'agit de définir des procédures d'agrégation (ou de vote) permettant de dégager une opinion collective à partir d'opinions individuelles. Depuis Condorcet (*cf.* Guilbaud [9]), de nombreuses voies de recherches ont été explorées. Toutes sont amenées à confronter leurs résultats au théorème d'Arrow [1]. La seconde approche correspond au problème de l'analyse de préférences. Il s'agit de définir une métrique sur l'ensemble des opinions permettant de mesurer des accords ou des divergences (*cf.* Kendall [11]). Le but est également d'obtenir une représentation la moins mauvaise possible, mais globale, de l'ensemble des proximités entre opinions, et éventuellement, entre opinions et objets de préférence (*cf.* Benzécri [4]).

Cet article expose une recherche plutôt orientée vers la seconde approche bien qu'il nous paraisse difficile de dissocier les deux problèmes.

Dans une première phase, nous avons précisé un modèle mathématique permettant de représenter une opinion. Nous avons choisi celui du tableau carré dont les éléments sont valués et soumis à des conditions de normalisation. Ce modèle est aussi celui du graphe valué correspondant à un tel tableau. Le fait de considérer une opinion collective comme valuée est courant. Généralement, le modèle de l'opinion individuelle est l'ordre (ordre total, partiel ou préordre) ou encore la relation de tournoi qui permet de représenter les intransitivités qui peuvent apparaître dans une opinion obtenue sous la forme de comparaisons par paires. Ces relations sont généralement non valuées (*cf.* Barbut et Monjardet [3]). Nous avons étendu la possibilité de valuer l'opinion individuelle pour deux raisons. D'une part, nous pensons que le même modèle doit pouvoir représenter l'opinion individuelle et l'opinion collective, à des conditions de normalisation près. D'autre part, des auteurs abordant le problème de choix à critères multiples (*cf.* Roy [16]) sont conduits à considérer l'opinion individuelle comme le résultat d'une agrégation d'opinions définies suivant les différents critères. On retrouve également cette thèse dans certains travaux expérimentaux tels ceux de Parlebas [15].

---

1. Je tiens à remercier MM. Barbut, Kreweras, Monjardet et Rosenstiehl pour leurs remarques au cours de la rédaction de cet article.

2. U.E.R. Mathématiques, Logique Formelle et Informatique, Paris V, I.S.H.A.

A partir de ce modèle de l'opinion individuelle et collective, nous avons défini plusieurs paramètres servant à caractériser de façon plus synthétique une opinion, tels les paramètres : intensité, cohérence, cohésion, puissance. Nous avons montré que certains de ces paramètres pouvaient être reliés à d'autres éléments connus dans le cas particulier de relations non valuées (distance et tau de Kendall).

Ces paramètres ont servi à déterminer des algorithmes d'agrégation d'opinions (ou de sujets). Leurs critères sont plus sociologiques que mathématiques. Ces algorithmes voudraient poser le problème de l'agrégation de sujets ou de formation de groupes en tant que processus dynamique.

Enfin une représentation géométrique simultanée des opinions (sujets) et des objets de préférence est proposée. Elle consiste en une projection sphérique sur le plan déterminé par deux axes privilégiés obtenus à l'aide des algorithmes, de l'ensemble des sujets et des objets.

I. OPINION INDIVIDUELLE

1. Définitions

Soit un ensemble fini I d'objets, d'options, de personnes, etc., sur lequel un ensemble A de sujets émet des opinions. Appelons opinion individuelle sur I le tableau carré privé de sa diagonale tel que :

$$\begin{cases} i, j \in I \\ a_{ij} \in [0, 1] \text{ (} [0, 1] \text{ est l'intervalle des nombres compris entre } 0 \text{ et } 1) \\ a_{ij} + a_{ji} \leqslant 1 \end{cases}$$

où :

a_{ij} est l'intensité de préférence : i préféré à j

a_{ji} est l'intensité de préférence : j préféré à i.

Dans l'introduction nous avons également parlé de graphe de préférence; en effet, à ce tableau correspond un graphe dont I est l'ensemble des sommets et où a_{ij} est la valuation de l'arc $i \to j$. Un graphe non valué sera représenté par un tableau semblable mais avec la condition plus restrictive :

$$a_{ij} \in \{0, 1\} \text{ au lieu de } a_{ij} \in [0, 1].$$

Posons : $n = \text{card } I$.

Il y a $n\,!$ façons de représenter une opinion individuelle, et cela en envisageant les $n\,!$ permutations des n objets qui constituent les indices des lignes et des colonnes du tableau.

Appelons base une des $n\,!$ permutations des objets ou, ce qui revient au même, une des permutations des indices de ligne et de colonne.

2. Cohérence

Appelons cohérence de l'opinion sur la base o la quantité :

$$\boxed{c_o = \frac{\sum\limits_{i < j, o} (a_{ij} - a_{ji})}{\sum\limits_{i < j, o} (a_{ij} + a_{ji})}}$$

o désignant l'ordre des lignes et des colonnes.

$\sum_{i<j,o} a_{ij}$ désigne la somme des termes du tableau situés au-dessus et à droite de la diagonale.

Lorsqu'on ne précisera pas la base sur laquelle la cohérence est définie, il s'agira d'un ordre total o à distance minimum de la relation binaire valuée a_{ij} envisagée, c'est-à-dire satisfaisant à :

$$\frac{\sum_{i<j} (a_{ij} - a_{ji})}{\sum_{i<j} (a_{ij} + a_{ji})} \text{ maximum.}$$

Pour l'expression de cette distance, voir III.3, et pour la recherche de o, voir Barbut [2] et Jacquet-Lagrèze [10].

Dans ce cas, on a :

$$0 \leqslant c \leqslant 1$$

$c = 1$, si il existe o, tel que $a_{ji} = 0$ quelque soit $i < j$, c'est-à-dire s'il n'y a pas de circuits dans le graphe de préférence.

3. *Intensité*

Appelons intensité la quantité :

$$\boxed{d = \frac{\sum_{i<j} (a_{ij} + a_{ji})}{n\,(n - 1)/2}}$$

On a :

$$0 \leqslant d \leqslant 1$$

car :

$$a_{ij} + a_{ji} \leqslant 1 \qquad \forall i, j \in I$$

d est indépendante de la base sur laquelle est exprimée l'opinion,
$d = 0$ correspond à une opinion vide. Le sujet ne donne aucune information sur les objets :

$$a_{ij} = 0, \forall i, j \in I$$

$d = 1$ correspond à une opinion « saturée », c'est-à-dire lorsque la relation binaire valuée est « complète » ou « totale », dans ce cas :

$$a_{ij} + a_{ji} = 1 \qquad \forall i, j \in I$$

4. *Puissance*

Appelons puissance de l'opinion sur la base o, la quantité :

$$p_o = \frac{\sum_{i<j,o} (a_{ij} - a_{ji})}{n\,(n - 1)/2},$$

on a :

$$0 \leqslant p \leqslant 1,$$

on a aussi la relation :

$$p_o = c_o\, d.$$

En effet :

$$p_o = \frac{\sum\limits_{i<j,o} (a_{ij} - a_{ji})}{\sum\limits_{i<j,o} (a_{ij} + a_{ji})} \times \frac{\sum\limits_{i<j,o} (a_{ij} + a_{ji})}{n\,(n-1)/2}\,.$$

Lorsqu'on ne précisera pas la base il s'agira de la base qui maximise la puissance (ou la cohérence et on la notera p ($p = c \cdot d$).

5. Opinions individuelles particulières

— Ordre total

L'ordre total est représenté par :

$$\begin{cases} a_{ij} \in \{0, 1\} \\ a_{ij} + a_{ji} = 1 \end{cases}$$
condition de transitivité

$$\begin{cases} a_{ij} = 1 \\ a_{ji} = 0 \text{ si } i \text{ est préféré à } j \end{cases}$$

Si on représente l'ordre total sur sa propre base, alors le tableau se présente avec des *1* dans le triangle supérieur et des *0* dans le triangle inférieur.

On a : $\forall i, j$ tel que $i < j$

$$\begin{cases} a_{ij} = 1 \\ a_{ji} = 0 \end{cases}$$

donc : $c = 1$, $d = 1$ et $p = 1$.

— Opinion sous la forme d'un graphe de tournoi

(Sur ces questions, voir Moon [14].)

Ce cas se présente lorsqu'on demande au sujet des comparaisons par paires ou plus généralement par blocs. Un bloc est un sous-ensemble de I sur lequel on demande d'établir un ordre total, et en constituant ces blocs de façon équilibrée, on obtient toutes les comparaisons 2 à 2 des éléments de I. En particulier, une paire est un bloc de 2 éléments. Un ordre total sur les n objets de I est un bloc de n éléments. Sur un tel exemple de questionnaires, voir les travaux de Huteau (I.N.O.P.) non encore publiés, sur des blocs de 4 éléments pour n (= 25) objets.

Le graphe de tournoi est représenté par :

$$\begin{cases} a_{ij} \in \{0, 1\} \\ a_{ij} + a_{ji} = 1 \end{cases} \qquad \begin{cases} a_{ij} = 1 \text{ si } i \text{ est préféré à } j. \\ a_{ji} = 0 \end{cases}$$

La seule différence avec l'ordre total est que l'on n'a plus la condition de transitivité.

On a alors :

$$c < 1, \quad d = 1, \quad p < 1 \qquad (p = c).$$

Problème

Trouver les graphes de tournoi les plus « défavorables » possibles, c'est-à-dire tels que c soit le plus faible possible.

Nous donnons les solutions pour n faible, le problème semble ouvert pour n quelconque.

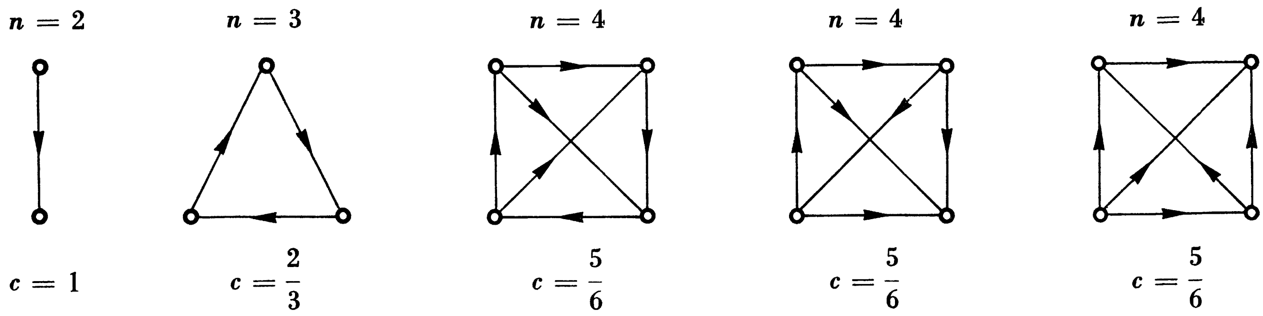
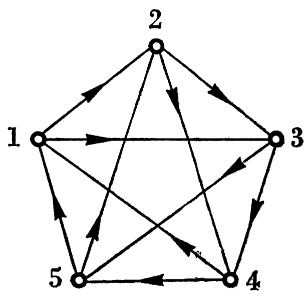


Fig. 1

	1	2	3	4	5
1	*0*	1	1	*0*	*0*
2	*0*	*0*	1	1	*0*
3	*0*	*0*	*0*	1	1
4	1	*0*	*0*	*0*	1
5	1	1	*0*	*0*	*0*

$n = 5$



$$c = \frac{7}{10}$$

Fig. 2

— Ordre partiel

Si i et j ne sont pas comparés (information incomplète) alors on représentera la paire (i, j) par :

$$\begin{cases} a_{ij} = 0 \\ a_{ji} = 0. \end{cases}$$

Le graphe d'un ordre partiel est sans circuit (transitivité) donc :

$$c = 1 \qquad d < 1 \qquad p < 1 \qquad (p = d).$$

Si l'ordre partiel est vide :

$$a_{ij} + a_{ji} = 0 \quad \forall i, j \Rightarrow d = p = 0.$$

— Préordre total

Considérons la paire (i, j) où i est jugé équivalent à j, on la représentera par

$$\begin{cases} a_{ij} = 1/2 \\ a_{ji} = 1/2. \end{cases}$$

Si on exprime le préordre total sur une base o à distance minimum, on aura :

$$c < 1 \text{ (présence de circuits) } d = 1 \qquad p = c.$$

II. OPINION COLLECTIVE

1. *Définition*

Soient A un ensemble de sujets, n_A son effectif.

Si $k \in A$ et si $\{a_{ijk}\}$ est l'opinion individuelle k, appelons opinion collective le tableau carré $n \times n$ défini par :

$$\boxed{A_{ij} = \sum_{k \in A} c_k \, a_{ijk}}$$

c_k est le poids de l'opinion individuelle k.

S'il s'agit d'opinions où tous les sujets ont même poids, alors c_k pourra être le nombre de sujets qui ont donné l'opinion d'indice k, et alors :

$$\sum_{k \in A} c_k = n_A = \textbf{card } A$$

2. *Cohésion d'un groupe*

Supposons l'opinion collective exprimée sur une base o (une des $n\,!$ permutations des n objets). Appelons cohésion du groupe sur la base o, ou encore cohésion du groupe autour de la tendance o, la quantité :

$$\boxed{C_{A,o} = \frac{\sum_{i < j, o} (A_{ij} - A_{ji})}{\sum_{i < j, o} (A_{ij} + A_{ji})}}$$

Lorsqu'on ne précisera pas la tendance, il s'agira de la tendance « centrale », définie comme étant l'ordre total o à distance minimum du tableau A_{ij}, ou encore o, tel que $C_{A,o}$ soit maximum.

En désignant cet ordre, on parlera aussi de la tendance du groupe A. Dans ce dernier cas et comme pour la cohérence de l'opinion individuelle, la cohésion est telle que :

$$0 \leqslant C_{A,o} \leqslant 1.$$

Cet indice de cohésion mesure une dispersion des opinions individuelles autour de la tendance o.

3. Intensité

Appelons intensité de l'opinion collective la quantité :

$$\boxed{D_A = \frac{\sum\limits_{i<j} (A_{ij} + A_{ji})}{n_A \cdot n\,(n-1)/2}}$$

Relation entre les intensités des opinions individuelles et l'intensité de l'opinion collective :

$$D_A = \frac{\sum\limits_{i<j} (A_{ij} + A_{ji})}{n_A \cdot n(n-1)/2} = \frac{\sum\limits_{i,j} A_{ij}}{n_A \cdot n(n-1)/2} = \frac{\sum\limits_{i,j} \sum\limits_{k \in A} c_k\, a_{ijk}}{n_A \cdot n(n-1)/2}$$

$$= \frac{\sum\limits_{k \in A} c_k \sum\limits_{i,j} a_{ijk}}{n_A \cdot n(n-1)/2}$$

or :

$$d_k = \frac{\sum\limits_{i,j} \cdot\, a_{ijk}}{n(n-1)/2}$$

donc :

$$D_A = \frac{\sum\limits_{k \in A} c_k\, d_k}{n_A}$$

Remarque

On a $D_A \leqslant 1$, $D_A = 1$ si tous les sujets donnent comme opinion une relation « saturée » (ex. : ordre ou préordre total, graphe de tournoi).

4. Puissance d'une opinion collective ou puissance d'un groupe

Appelons puissance d'un groupe A autour de la tendance o la quantité :

$$\boxed{P_{A,o} = \frac{\sum\limits_{i<j,o} (A_{ij} - A_{ji})}{n(n-1)/2}}$$

Lorsqu'on ne précisera pas la tendance o, il s'agira d'un ordre total o à distance minimum de l'opinion, donc o tel que P_A soit maximum.

Relation entre P_A, C_A et D_A :

On montre facilement que :

$$P_A = n_A\, C_A\, D_A.$$

Remarque

— $C_A \leqslant 1$, $D_A \leqslant 1$ donc $P_A \leqslant n_A$,

— $P_A = n_A \Leftrightarrow$ tous les sujets ont pour opinion le même ordre total.

Exemple 1

Soit le système d'opinions suivant :
un groupe G_1 de 30 sujets donnant tous l'ordre $1 > 2 > 3$ ($>$ désignant préféré à),
un groupe G_2 de 30 sujets donnant tous l'ordre $3 > 2 > 1$.

Nous avons : $P_{G_1} = 30$ $\quad P_{G_2} = 30$ $\quad$ et $\quad P_{G_1 \cup G_2} = 0$.

En effet :

$$G_1 \cup G_2 : \quad \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & 0 & 30 & 30 \\ 2 & 30 & 0 & 30 \\ 3 & 30 & 30 & 0 \end{array}$$

$C_{G_1 \cup G_2} = 0$ $\quad$ et $\quad$ $P_{G_1 \cup G_2} = 0$.

Exemple 2

Soient n objets et $n\,!$ sujets. Chaque sujet donne comme opinion une des $n\,!$ permutations des n objets, 2 sujets ne donnant jamais la même. L'opinion collective est alors :

$$\begin{array}{c|cccc} & 1 & 2 & \dots & n \\ \hline 1 & 0 & n\,!/2 & & n\,!/2 \\ 2 & n\,!/2 & 0 & & n\,!/2 \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ n & n\,!/2 & n\,!/2 & & 0 \end{array} \qquad \text{et} \qquad \begin{array}{l} C_A = 0 \\ D_A = 1 \\ \\ P_A = 0 \end{array}$$

III. RELATIONS ENTRE DISTANCES ET PUISSANCES

1. *Compléments sur les distances*

Soient 2 opinions individuelles $\{k_{ij}\}$ et $\{l_{ij}\}$, nous avons pris pour distance la quantité :

$$d\,(k, l) = \sum_{i,j} | k_{ij} - l_{ij} |$$

— La distance maximum est $d\,(k, l) = n\,(n - 1)$, nous la noterons $dmax$.

— Lorsque k_{ij}, $l_{ij} \in \{0, 1\}$, les opinions individuelles sont des relations binaires non valuées et cette distance est le double de la distance d_K de Kendall définie par :

$$d_K\,(k, l) = \frac{1}{2} \sum_{i,j} | k_{ij} - l_{ij} | = \frac{1}{2} \sum_{i,j} (k_{ij} - l_{ij})^2$$

cf. [11], [5].

2. Distance entre 2 opinions collectives

Soient deux groupes distincts : A, n_A, $\{A_{ij}\}$ et B, n_B, $\{B_{ij}\}$,

on posera :

$$d\,(A, B) = \sum_{i,j} |\, A_{ij}/n_A - B_{ij}/n_B \,|$$

On vérifie que cette quantité répond bien aux axiomes de la distance.

3. Distance entre une opinion collective et sa tendance

Nous avons vu que la tendance d'une opinion collective est l'ordre total à distance minimum de cette opinion collective.

Appelons $\{k_{Ai,j}\}$ la tendance du groupe A, $\{A_{ij}\}$:

$$k_{Ai,j} \in \{0, 1\}.$$

La distance entre l'opinion collective et sa tendance est :

$$d\,(A, k_A) = \sum_{i,j} |\, A_{ij}/n_A - k_{Ai,j} \,|.$$

On peut montrer (*cf.* [10]) que :

$$d\,(A, k_A) = \frac{1}{n_A} \sum_{k \in A} d(k, k_A).$$

4. Autres expressions de la puissance

Évaluons la quantité $\sum_{k \in A} d(k, k_A)$ où k_A est la tendance de A :

$$k_{Aij} \in \{0, 1\}, k_{ij} \in [0, 1]$$

$$\sum_{k \in A} d(k, k_A) = \sum_{k \in A} \sum_{i,j} |k_{ij} - k_{Aij}\,|.$$

Nous passerons des intermédiaires obtenus, en considérant les cas où :

$$k_{Aij} = 1 \qquad \text{et} \qquad k_{Aij} = 0.$$

On montre que si on choisit l'ordre des indices i dans la sommation $\sum_{i<j}$, égal à l'ordre de la tendance k_A, alors cette expression devient :

$$\sum_{k \in A} d(k, k_A) = \sum_{k \in A} \sum_{i<j,\, k_A} (1 - (k_{ij} - k_{ji}))$$

soit :

$$= \sum_{k \in A} \left[\frac{n(n-1)}{2} - \sum_{i<j,\, k_A} (k_{ij} - k_{ji}) \right]$$

$$= \frac{n(n-1)}{2}\, n_A - \frac{n(n-1)}{2}\, P_A$$

ce qui donne si $dmax = n(n-1)$, et en tenant compte de l'expression de $d(A, k_A)$, les trois formules suivantes :

$P_A = n_A - \sum_{k \in A} \frac{d(k, k_A)}{dmax/2}$	$P_A = n_A \left(1 - \frac{d(A, k_A)}{dmax/2}\right)$
$P_A = \sum_{k \in A} \left(1 - \frac{d(k, k_A)}{dmax/2}\right)$	

Relation avec le tau de Kendall τ

Le tau de Kendall entre 2 opinions α et β est :

$$\tau_{\alpha\beta} = 1 - 2\,\frac{d_K(\alpha,\beta)}{n(n-1)/2}$$

Comme la distance que nous envisageons est le double de celle de Kendall, nous avons :

$$1 - \frac{d(k, k_A)}{dmax/2} = \tau_{k,k_A}$$

soit :

$$\boxed{P_A = \sum_{k \in A} \tau_{k,k_A}}$$

où k_A est la tendance du groupe A.

5. Formule barycentrique

Soient les groupes disjoints A, n_A, $\{A_{ij}\}$ et B, n_B, $\{B_{ij}\}$, on a :

$$d(A, B) = \sum_{i,j} | A_{ij}/n_A - B_{ij}/n_B |.$$

En considérant le groupe formé par la réunion des 2, noté $A + B$, d'effectif $n_A + n_B$, et d'opinion $\{A_{ij} + B_{ij}\}$, on montre facilement les relations :

$d(A, A+B) = \frac{n_B}{n_A + n_B}\, d(A, B)$	$\frac{d(A, B)}{n_A + n_B} = \frac{d(A, A+B)}{n_B} = \frac{d(B, A+B)}{n_A}$
$d(B, A+B) = \frac{n_A}{n_A + n_B}\, d(A, B)$	$d(A, B) = d(A, A+B) + d(B, A+B)$

6. Composition des puissances

Le problème est de déterminer les puissances d'un groupe obtenu par la réunion de 2 groupes disjoints, à partir d' éléments« résumant » les caractéristiques de ces groupes telles que leur effectif, leur puissance, leur tendance, etc.

Cas simplifié

Soient 2 groupes A et B disjoints de cohésion 1, n_A et n_B, leurs effectifs, k_A et k_B leurs tendances.

Cherchons à évaluer la puissance P_{A+B} du groupe réunissant les 2 groupes A et B.

Si k_C représente la tendance du groupe résultant, on a :

$$P_{A+B} = \sum_{k \in A+B} \left(1 - \frac{d\,(k, k_C)}{dmax/2}\right)$$

$$= n_A + n_B - \sum_{k \in A} \frac{d\,(k, k_C)}{dmax/2} - \sum_{k \in B} \frac{d\,(k, k_C)}{dmax/2}.$$

Si : $k \in A, k = k_A$ et si : $k \in B, k = k_B$, donc :

$$P_{A+B} = n_A + n_B - 2/dmax\,[n_A\, d(k_A, k_C) + n_B\, d\,(k_B, k_C)].$$

En examinant les cas possibles pour $k_{A_{ij}}$, $k_{B_{ij}}$ et $k_{C_{ij}}$, suivant les valeurs relatives de n_A et n_B, on montre que :

$$n_A\, d\,(k_A, k_C) + n_B\, d\,(k_B, k_C) = d\,(k_A, k_B) \inf\,(n_A, n_B).$$

On alors une relation de composition des puissances dans ce cas particulier :

$$\boxed{P_{A+B} = n_A + n_B - 2\,\frac{d\,(k_A, k_B)}{dmax} \inf\,(n_A, n_B).}$$

Cas général

P_A, P_B sont quelconques.

$$P_A = P_{A,k_A} = n_A\left(1 - \frac{d\,(A, k_A)}{dmax/2}\right).$$

La puissance de A sur une base quelconque o est :

$$P_{A,o} = n_A\left(1 - \frac{d\,(A, o)}{dmax/2}\right),$$

d'où une formule analogue au théorème de Huygens pour les moments d'inertie :

$$\boxed{P_{A,o} = P_A - \frac{n_A}{dmax/2}\,[d\,(A, o) - d\,(A, k_A)]}$$

Exprimons les puissances des groupes disjoints A et B sur la tendance k_C de leur réunion :

$$P_{A,k_c} = P_A - \frac{n_A}{dmax/2}\,[d\,(A, k_C) - d\,(A, k_A)]$$

$$P_{A,k_c} = P_B - \frac{n_B}{dmax/2}\,[d\,(B, k_C) - d\,(B, k_B)]$$

d'où, en faisant la somme :

$$P_{A,k_c} + P_{B,k_c} = P_A + P_B - \frac{2}{dmax}\left[n_A\, d(A, k_C) + n_B\, d(B, k_C) - n_A\, d(A, k_A) - n_B\, \mathrm{d}(B, k_B)\right].$$

Or le premier membre représente la puissance du groupe $A + B$ sur sa tendance k_C, donc, la puissance du groupe résultant :

$$P_{A+B} = P_A + P_B - \frac{2}{dmax}\left[n_A,\, d(A, k_C) + n_B\, d(B, k_C) - n_A\, d(A, k_A) - n_B\, d(B, k_B)\right].$$

Il semble que dans certains cas, cette formule puisse se réduire à :

$$P_{A+B} = P_A + P_B - 2\,\frac{d(k_A, k_B)}{dmax}\,\inf(n_A, n_B),$$

mais dans d'autres cas, elle n'est pas vérifiée.

IV. ALGORITHMES D'AGRÉGATION

1. *Agrégation artificielle*

Étant donné un système d'opinions sur des objets, on voudrait construire le, ou éventuellement, les ordres totaux sur les n objets à distance minimum de l'opinion collective $\{A_{ij}\}$.

On peut alors qualifier cet ordre total d'opinion collective à condition de lui associer certaines grandeurs telles la cohésion, l'intensité, la puissance. Cette démarche est analogue à celle qui consiste à résumer un ensemble de données par une moyenne et une variance.

Il peut cependant être intéressant d'établir des partitions sur l'ensemble des sujets, en s'appuyant sur la notion de puissance.

L'intérêt de cette notion est qu'elle permet de mettre en évidence des tensions entre sujets et entre groupes, des formations de groupes, etc. Dans l'exemple 1 du § II. 4, les deux groupes de sujets G_1 et G_2 ne chercheront pas forcément à s'agréger en un seul groupe $G_1 + G_2$ qui serait incapable de se décider, car incapable d'émettre une préférence marquée pour 1, 2 ou 3.

2. *L'agrégation « spontanée »*

Nous désignerons par ce terme des modes d'agrégation de sujets qui conduisent à donner une partition de ces sujets.

L'expérience sur des données réelles a montré que lorsqu'on étudie un système d'opinions, on obtient souvent une opinion collective de cohésion faible donc de puissance faible.

Si on prend un système d'opinions dont chaque opinion individuelle est un ordre total, la puissance de chaque « groupe » de 1 individu est égale à 1. Les sujets d'opinions voisines accepteront de se regrouper de façon à constituer un groupe plus puissant, même si la tendance de ce groupe n'est pas tout à fait l'opinion de chaque sujet.

Si on considère une agrégation des sujets représentée sous la forme d'une arborescence (*cf.* Benzécri [4] et Lerman [13], et si l'on portait à chaque nœud, la puissance du groupe ainsi constitué, on verrait que dans certains cas, et de façons bien différentes, il est possible de constituer des groupes dont la puissance soit supérieure à celle de l'ensemble des sujets.

Groupes en tension

On dira que deux groupes disjoints sont en tension lorsque :

$$P_{A+B} < \sup (P_A, P_B).$$

Dans le cas contraire, on dira qu'ils s'agrègent spontanément.
Cas particulier de deux individus.

On a vu la relation :

$$P_{A+B} = n_A + n_B - 2 \frac{d(k_A, k_B)}{dmax} \inf (n_A, n_B),$$

pour 2 sujets k, l;

$$n_A = n_B = 1, \qquad k_A = k \qquad \text{et} \qquad k_B = l,$$

alors :

$$P_{k+l} = 2 - 2 \frac{d(k, l)}{dmax} = 2\left(1 - \frac{d(k, l)}{dmax}\right),$$

par conséquent, deux individus sont en tension si $P_{k+l} < 1$:

$$\boxed{\begin{array}{l} d(k, l) > dmax/2 \Leftrightarrow k \text{ et } l \text{ sont en tension} \\ \qquad\qquad\qquad\quad \Leftrightarrow \\ \tau_{kl} < 0 \end{array}}$$

τ_{kl} étant le tau de Kendall.

Groupe en équilibre autour d'une tendance

On dira qu'un groupe A est en équilibre autour d'une tendance k_A si tous les sujets du groupe contribuent à augmenter la puissance de A autour de k_A.

Théorème

Un groupe A est en équilibre autour d'une tendance k_A si et seulement si toutes les opinions k du groupe sont à une distance de k_A inférieure ou égale à $dmax/2$.

Cette proposition résulte de la formule :

$$P_{A, k_A} = \sum_{k \in A} \left(1 - \frac{d(k, k_A)}{dmax/2}\right).$$

Si $d(k, k_A) > dmax/2$, k contribue à affaiblir P_{A, k_A}.

Si $d(k, k_A) < dmax/2$, k contribue à augmenter P_{A, k_A}.

Autre énoncé : Un groupe est en équilibre autour d'une tendance k_A si et seulement si, les « tau » de Kendall entre les opinions des éléments du groupe et la tendance k_A sont tous positifs ou nuls.

Ceci résulte de $P_A = \sum_{k \in A} \tau_{k, k_A}$.

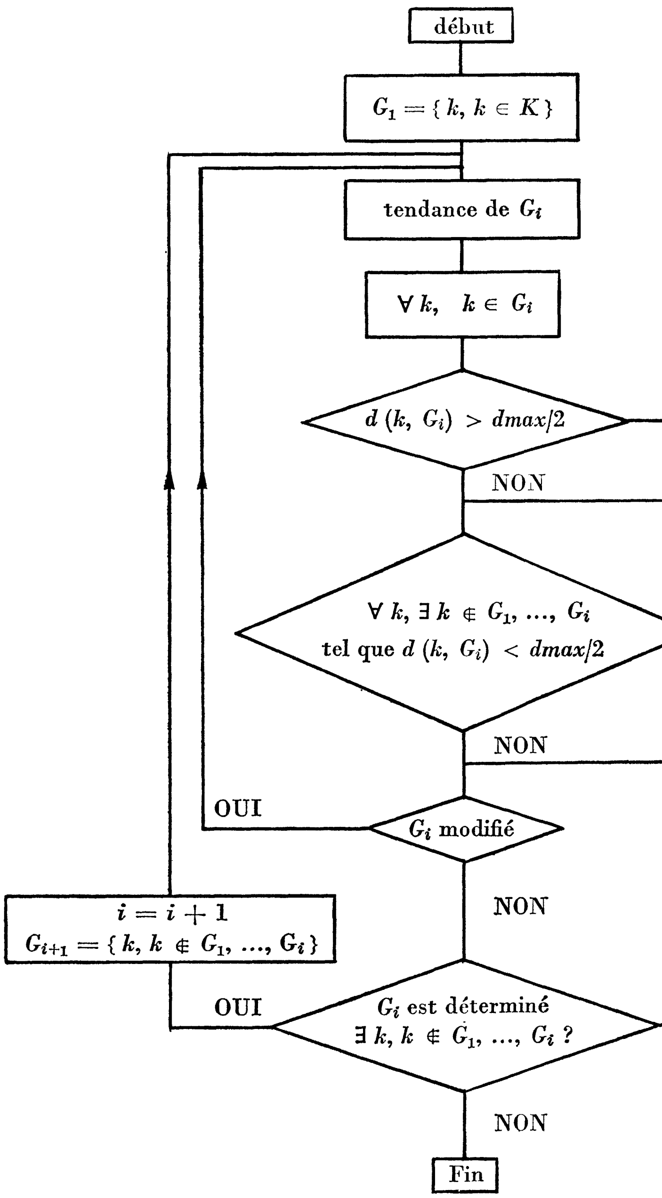
3. Exemples d'algorithmes

Nous présenterons deux algorithmes s'appuyant sur ces notions, le second pouvant être la suite du premier.

Algorithme 1

Il répond au critère d'agrégation : se rattacher au groupe de la tendance la plus centrale à condition de ne pas être rejeté, ou encore s'agréger « si possible » à la majorité.

On notera : $G_1, \ldots, G_i$ les différents groupes constitués, $K = \{k\}$ l'ensemble des sujets et $d\,(k, G_i)$ représentera la distance de k à la tendance du groupe G_i.

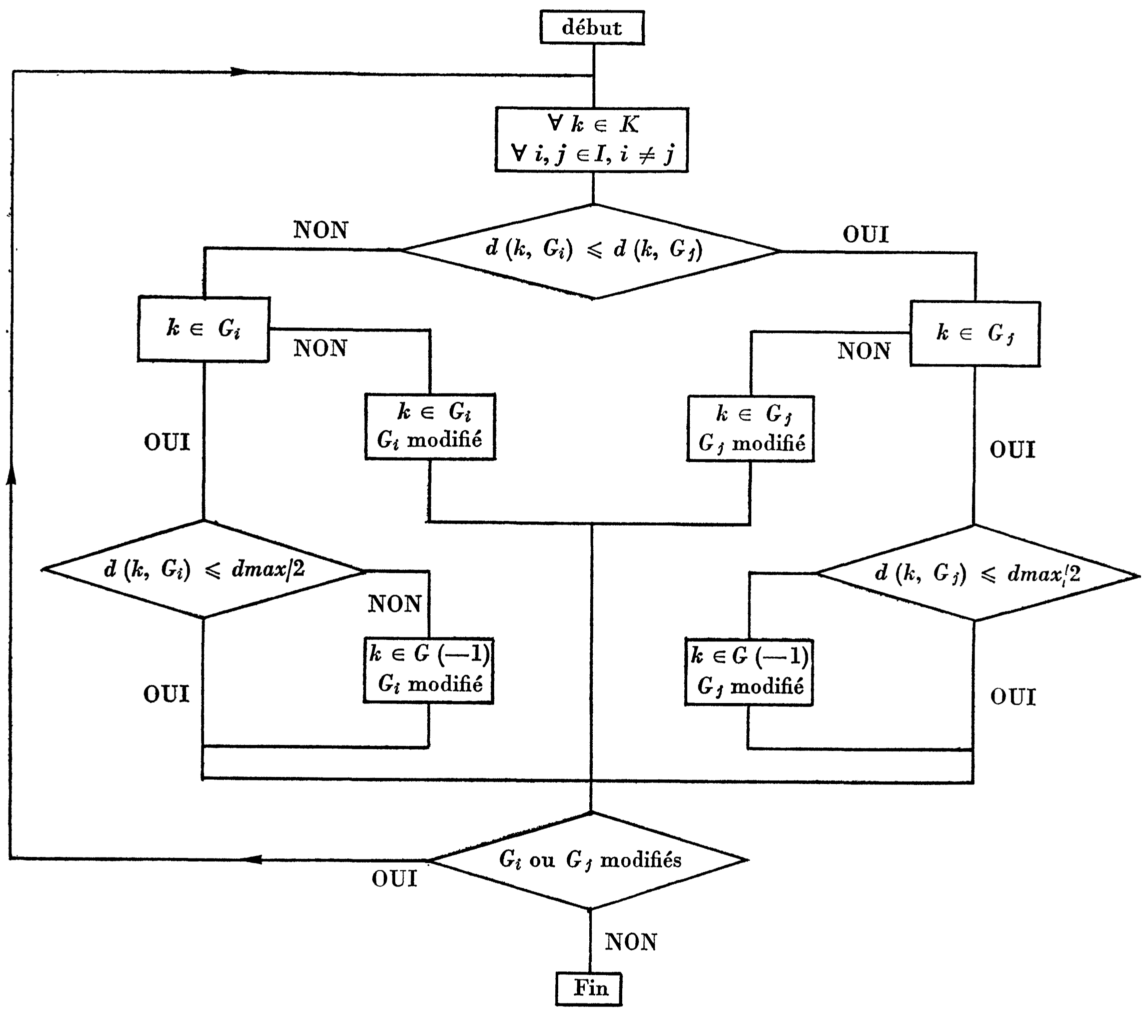


Algorithme 2

Il est la suite de l'algorithme 1, lorsqu'au moins 2 groupes se sont formés.

Il répond au critère d'agrégation : s'agréger au groupe dont la tendance est la plus voisine de son opinion.

Soit I l'ensemble des indices des groupes G_i, $i \in I$.



$G(-1)$ est un groupe qui reçoit des sujets expulsés des groupes au cours de l'algorithme par suite des déplacements des tendances.

Remarque

Ces algorithmes ont été programmés et insérés dans un programme d'analyse de préférences. Sur les données traitées, l'algorithme 1 a conduit à 1, 2 ou 3 groupes, dont l'un d'entre eux est plus important en effectif, et dont la tendance est généralement voisine de celle de l'ensemble des sujets (le groupe majorité).

L'algorithme 2 conduit généralement à 2 groupes plus équilibrés en effectif. Parfois un troisième groupe apparaît [le groupe G (— 1)].

Groupes stables

On a vu que dans un groupe en équilibre autour d'une tendance, aucun des sujets n'est en tension avec la tendance centrale.

Cependant, il se peut très bien que 2 sujets soient en tension entre eux. En effet, des groupes en équilibre autour d'une tendance peuvent être instables. Si certains sujets quittent le groupe, la tendance se déplace, et certains sujets du groupe peuvent être alors en tension avec la nouvelle tendance.

Le départ de quelques sujets peut provoquer la désagrégation partielle d'un groupe en équilibre autour d'une tendance.

Une autre notion est alors intéressante : celle de groupe stable.

Définition

Appelons groupe stable un groupe dans lequel tous les sujets ne sont pas en tension 2 à 2, ou encore : G est un groupe stable :

$$\Leftrightarrow \forall\, k, l \in G, \quad d\,(k, l \leqslant \frac{dmax}{2}\,.$$

Trouver des groupes stables, c'est trouver des cliques dans le graphe :

$$\exists \text{ arête } k, l \text{ si } d\,(k, l) \leqslant \frac{dmax}{2}\,.$$

Remarque

Un groupe stable est en équilibre autour de la tendance de tout sous-groupe, donc en particulier autour de toute opinion individuelle de ce groupe.

V. REPRÉSENTATION GÉOMÉTRIQUE

Nous définirons le permutoèdre de la façon suivante : Angle sous lequel on voit deux opinions (ordres totaux) sur le permutoèdre :

$$\boxed{\alpha_{kl} = \pi \cdot \frac{d\,(k, l)}{dmax}} \qquad \alpha \text{ en radians}$$

$$\alpha_{kl} = 180 \cdot \frac{d\,(k, l)}{dmax} \qquad \alpha \text{ en degrés}$$

Lien avec le tau de Kendall

$$\tau_{kl} = 1 - 2\,\frac{d\,(k, l)}{dmax}$$

donc :

$$\boxed{\alpha_{kl} = \frac{\pi}{2}\,(1 - \tau_{kl})}$$

Il est donc possible de représenter un ensemble d'opinions $\{k\} = A$ par rapport à deux axes, qui sont deux tendances G_1, G_2.

En effet, on peut définir deux angles : α (k, G_1) ou plus simplement $(k\ G_1)$ et α (k, G_2), $(k\ G_2)$ et le rayon $Ok = \rho = d_k$ (intensité de k).

On peut ensuite envisager une projection sphérique des sujets sur les deux axes G_1 et G_2, suivant les formules que nous allons établir dans le paragraphe suivant.

Formules de projection

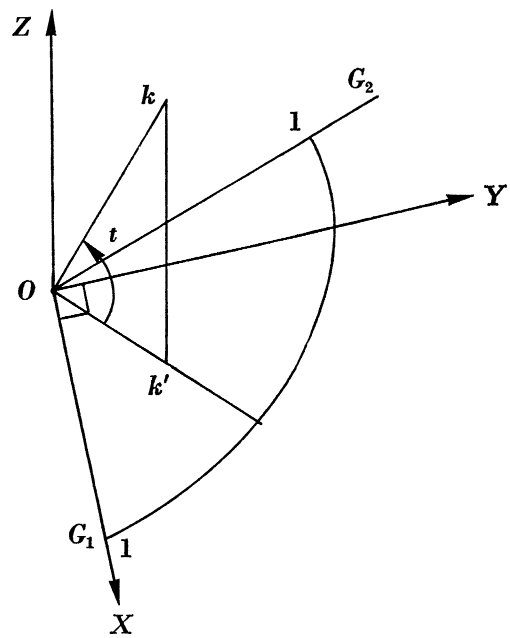


Fig. 3

On définit les axes X et Y de la façon suivante :

pour $G_1 : X = 1 \qquad Y = 0$

pour $G_2 : X = \cos\,(G_1\ G_2) \qquad Y = \sin\,(G_1\ G_2)$

en désignant par $G_1\ G_2$ l'angle α (G_1, G_2).

On cherchera à évaluer les coordonnées $X_k{}'$, et $Y_k{}'$, du point k' projection orthogonale du point k sur le plan X, Y.

$$X_k{}' = \rho \cos t \cos\,(k'\ G_1) \qquad \text{si} \qquad t = (Ok', Ok)$$

$$Y_k{}' = \rho \cos t \sin\,(G_1, k') = -\,\rho \cos t \sin\,(k'\ G_1).$$

On a les relations :

$$\vec{k} \cdot \vec{G_1} = \vec{k'} \cdot \vec{G_1} + \vec{Z} \cdot \vec{G_1}$$

$$= \vec{k'} \cdot \vec{G_1}$$

4

Or $|k'| = \cos t \cdot \rho$ donc :

$$\rho \cos (k\ G_1) = \rho \cos t \cos k'\ G_1.$$

Et de même :

$$\vec{k} \cdot \vec{G_2} = \vec{k'} \cdot \vec{G_2} + \vec{Z} \cdot \vec{G_2}$$
$$= \vec{k'} \cdot \vec{G_2}$$

Soit :

$$\rho \cos (k\ G_2) = \cos t \cos k'\ G_2$$

soit :

$$\frac{\cos k'\ G_2}{\cos k'\ G_1} = \frac{\cos k\ G_2}{\cos k\ G_1}$$

De plus on a la relation :

$$k'\ G_2 = k'\ G_1 + G_1\ G_2$$

soit :

$$\cos k'\ G_2 = \cos k'\ G_1 \cos G_1\ G_2 - \sin k'\ G_1 \sin G_1\ G,$$

donc :

$$\frac{\cos k'\ G_2}{\cos k'\ G_1} = \cos G_1\ G_2 - \operatorname{tg} k'\ G_1 \sin G_1\ G_2$$

$$\frac{\cos k'\ G_2}{\cos k'\ G_1} = \frac{\cos k\ G_2}{\cos k\ G_1}$$

on en déduit :

$$\operatorname{tg} k'\ G_1 = \left[\cos G_1\ G_2 - \frac{\cos k\ G_2}{\cos k\ G_1}\right] \frac{1}{\sin G_1\ G_2}$$

On a donc :

$$X_{k'} = \rho \cos t \cos k'\ G_1 = \rho \cos k\ G_1$$
$$Y_{k'} = - \rho \cos t \sin k'\ G_1 = - \rho \cos k\ G_1 \operatorname{tg} k'\ G_1$$

soit :

$$Y_{k'} = - [\cos k\ G_1 \cos G_1\ G_2 - \cos k\ G_2] \frac{1}{\sin G_1\ G_2}$$

D'où les formules de projection de k sur le plan $G_1\ G_2$ connaissant $k\ G_1$ et $k\ G_2$:

$$\boxed{\begin{aligned} X_{k'} &= \rho \cos k\ G_1 \\ Y_{k'} &= \rho\ [\cos k\ G_2 - \cos k\ G_1 \cos G_1\ G_2]/\sin G_1\ G_2 \end{aligned}}$$

avec $\rho = d_k =$ intensité de l'opinion k.

Remarque

Lorsque :

$$k\ G_2 = k\ G_1 + G_1\ G_2,\ \cos k\ G_2 = \cos k\ G_1 \cos G_1\ G_2 - \sin k\ G_1 \sin G_1\ G_2$$

et :

$$Y_{k,} = -\rho \sin k\ G_1 = \rho \sin G_1\ k,$$

on retrouve k et k' confondus sur le cercle de rayon ρ.

PROJECTION DES OBJETS

1. Disposition des objets sur le permutoèdre

Soit l'objet X et $\alpha \ldots \mu$, les $n - 1$ autres objets. Pour situer les objets, nous chercherons à satisfaire la proposition suivante :

L'objet X doit occuper sur le permutoèdre une position symétrique par rapport au permutoèdre des $n - 1$ autres objets.

2. Définition de la distance d'un objet X à un ordre O_1

Supposons que rang $X = q < \frac{n}{2}$ et considérons les deux ordres :

$$O_1 : \alpha \ldots \delta\ X\ \varepsilon \ldots \xi\ \rho \ldots \mu$$
$$O_2 : \mu \ldots \rho\ X\ \xi \ldots \varepsilon\ \delta \ldots \alpha$$

tels que $O_1 - \{X\}$ et $O_2 - \{X\}$ soient deux permutations inverses, nous poserons par définition :

$$\boxed{d\,(O_1, X) = d\,(O_2, X) = \frac{1}{2}\, d\,(O_1, O_2)}$$

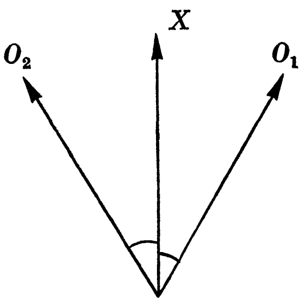


Fig. 4

Évaluons $d\,(O_1, O_2)$:

	1	$q - 1$	q	$q + 1$		n
O_1	$\alpha \ldots$	δ	X	ε	$\ldots \xi\ \rho \ldots$	μ
O_2	$\mu \ldots$	ρ	X	ξ	$\ldots \varepsilon\ \delta \ldots$	α

ξ et ε sont définis tels que : **card** $\{\alpha \dots \delta\}$ = **card** $\{\rho \dots \mu\}$

On a :

$$\textbf{card}\ (\varepsilon \dots \xi) = n - 1 - \textbf{card}\ (\alpha \dots \delta) - \textbf{card}\ (\rho \dots \mu)$$
$$= n - 1 - 2\,(q - 1)$$

$d\,(O_1, O_2)$ étant le nombre d'arcs en désaccord dans O_1 et O_2.

Le nombre d'arcs en accord est 2 **card** $(\varepsilon \dots \xi)$.

En effet, sur la base O_1, O_2 s'exprime par :

$$\begin{array}{c|c|c|c|c|} & \alpha \dots \delta & X & \varepsilon \dots \xi & \rho \dots \mu \\ \hline \begin{matrix}\alpha\\ \vdots \\ \delta\end{matrix} & \begin{matrix} o & 0 \\ 1 & o\end{matrix} & \begin{matrix} o \\ o \end{matrix} & 0 & 0 \\ \hline X & o \quad o & o & 1 \quad 1 & o \quad o \\ \hline \begin{matrix}\varepsilon\\ \vdots \\ \xi\end{matrix} & 1 & \begin{matrix} o\\ o\end{matrix} & \begin{matrix} o & 0 \\ 1 & o\end{matrix} & 0 \\ \hline \begin{matrix}\rho\\ \vdots \\ \mu\end{matrix} & 1 & \begin{matrix}1\\1\end{matrix} & 1 & \begin{matrix} o & \\ & o\end{matrix} \\ \hline \end{array}$$

$$\Rightarrow d\,(O_1, O_2) = n\,(n - 1) - 2\ \text{card}\ (\varepsilon \dots \xi)$$
$$= n\,(n - 1) - 2\,[n - 1 - 2\,(q - 1)]$$
$$= n\,(n - 1) - 2\,(n + 1 - 2\,q)$$

et :

$$d\,(X, O_1) = d\,(X, O_2) = \frac{1}{2}\,d\,(O_1, O_2)$$

Cette distance ne dépend que du rang de X sur O_i.

La relation rentre la distance et l'angle est :

$$\alpha = \frac{\pi}{n\,(n - 1)}\,d$$

donc l'angle d'un objet X du permutoèdre à un ordre total O_i (permutation de ce permutoèdre) est :

$$\alpha\,(X, O_i) = \frac{\pi}{n\,(n - 1)}\,\frac{1}{2}\,[n\,(n - 1) - 2\,(n + 1 - 2\,q)]$$

soit :

$$\boxed{\alpha(X, O_i) = \frac{\pi}{2} - \frac{\pi}{n(n-1)}(n + 1 - 2 \text{ Rang } (X, O_i))}$$

Cas où rang $(X, O_i) \geqslant \frac{n}{2}$

En posant :

$$\alpha(X, O_1) = \alpha(X, O_2) = \pi - \frac{1}{2}\alpha(O_1, O_2),$$

on démontre que la formule ci-dessus est encore valable.

Remarque 1

Il existe un programme (ANAPRE) qui exécute les algorithmes 1 et 2 à partir de données sous forme de préordres partiels (donc en particulier, d'ordres totaux). La suite du programme donne une projection graphique du système d'opinions sur les axes G_1 et G_2.

Remarque 2

Plusieurs systèmes réels d'opinions ont été analysés, le plus important portant sur 480 sujets et 25 objets de préférence.

Remarque 3

Les permutoèdres ($n = 4$) et ($n = 5$) ont été projetés de façon à connaître toutes les figures possibles suivant l'angle des deux axes de projection G_1 et G_2. On donne en annexe le permutoèdre ($n = 3$).

Remarque 4.

Dans cette dernière optique, on pourrait chercher à représenter toutes les relations non valuées, totales ou partielles, ou encore tous les graphes possibles orientés, sans boucles et antisymétriques.

Si on pose $n' = n(n-1)/2$ (n objets), le nombre de graphes de p arcs disposés entre n sommets est : $2^p \, C^p_{n'}$ (c'est encore le nombre d'opinions d'intensité p/n').

Le nombre d'opinions distinctes est alors :

$$2^0 C^0_{n'} + 2 C^1_{n'} + \dots + C^p_{n'} 2^p + \dots + C^n_n 2^{n'} = 3^{n'} = 3^{n(n-1)/2}$$

$2^0 C^0_{n'}$ pour l'opinion vide ($\varnothing$)

$C^{n'}_{n'} 2^{n'} = 2^{n(n-1)/2}$ est le nombre de relations totales à $n(n-1)/2$ arcs, ou encore le nombre d'opinions d'intensité 1.

Permutoèdre $N = 3$

Projection des ordres totaux (6)
Projection de certains ordres partiels

$$\text{axes} \begin{cases} G_1 = \text{BCA} \\ G_2 = \text{ABC} \end{cases}$$

Un trait gras représente une distance : 1.

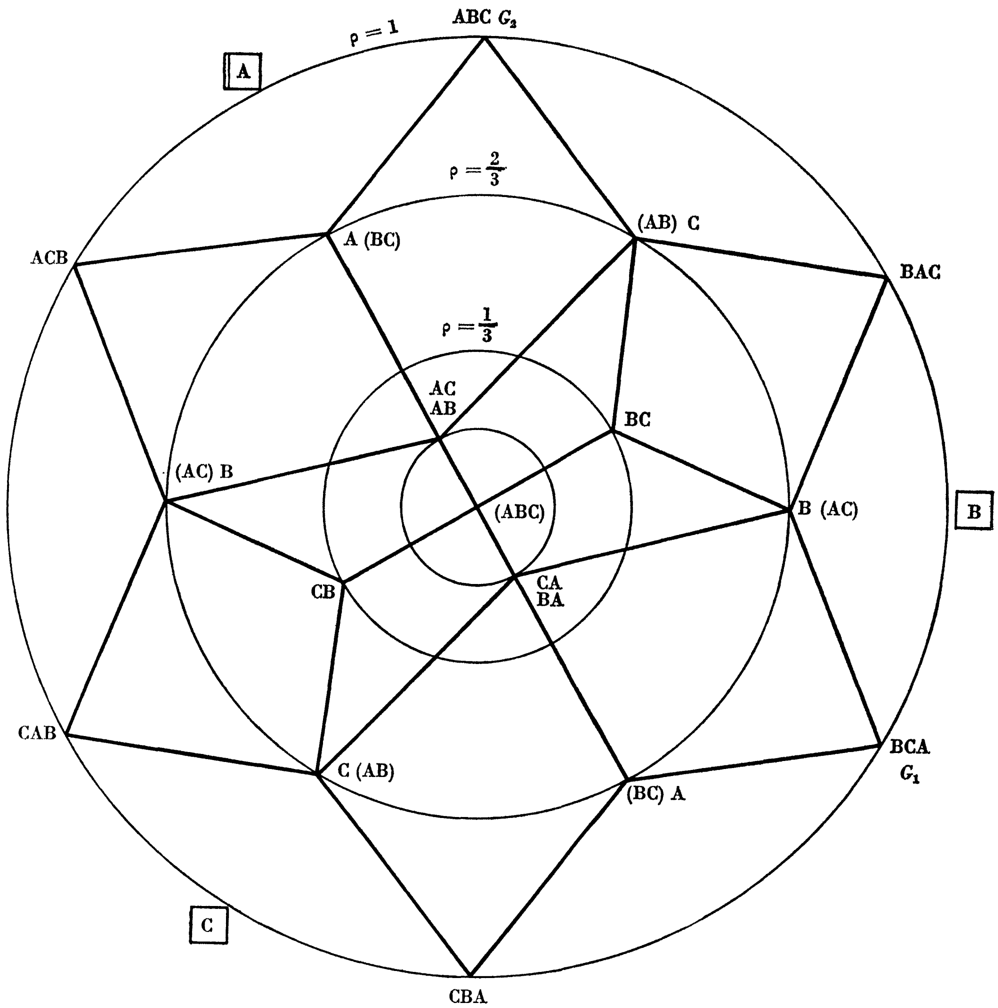


Fig. 5

BIBLIOGRAPHIE

[1] Arrow, K. J., *Social choice and individual values*, 2e éd., Wiley, 1962.

[2] Barbut, M., " Note sur les ordres totaux à distance minimum d'une relation binaire donnée", *Math. Sci. hum.*, n° 17.

[3] Barbut, M. et Monjardet, B., *Ordre et classification, algèbre et combinatoire*, 2, Paris, Hachette, 1970.

[4] Benzecri, J. P., *Sur l'analyse des préférences*, Paris, publication I.S.U.P. ; *L'analyse factorielle des correspondances*, Paris, publication I.S.U.P. ; *Classification automatique*, Paris, publication I.S.U.P.

[5] Degenne, A., *L'ajustement des modèles ordonnés en analyse des questionnaires*, Thèse, 1969.
A paraître : " Techniques ordinales en analyse des données ", *Métrique et statistique*, Paris, Hachette, 1971.

[6] Feldmann-Hogaasen, J., " Ordres partiels et permutoèdres ", *Math. Sci. hum.*, n° 28, 1969.

[7] Frey, L., " Techniques ordinales en analyse des données ", *Algèbre et Combinatoire* (A paraître), Paris, Hachette, 1971.

[8] Guilbaud, G. Th. et Rosenstiehl, P., " Analyse algébrique d'un scrutin ", *Math. Sci. hum.*, n° 4, 1963.

[9] Guilbaud, G. Th., " La théorie de l'intérêt général et le problème logique de l'agrégation ", *Monographies de Recherche Opérationnelle*, n° 9, Dunod, 1968.

[10] Jacquet-Lagrèze, E., " L'agrégation des opinions individuelles ", *Informatique en Sciences Humaines*, n° 4, 1969.

[11] Kendall, M. G., *Rank correlation methods*, Londres, Griffin, 1962.

[12] Kreweras, G., " Représentation polyédrique des préordres complets finis et application à l'agrégation des préférences ", *La décision*, 2, C.N.R.S., 1967.

[13] Lerman, I. C., *Les bases de la classification automatique*, Paris, Gauthier-Villars, 1970.

[14] Moon, J. W., *Topics on tournaments*, New York, Holt, 1968.

[15] Parlebas, P., *Effet condorcet et dynamique sociométrique*, Brochure ronéotée, U.E.R.M.L.I., Paris V, 1970.

[16] Roy, B., " Classements et choix en présence de points de vue multiples " (la méthode Électre), *Rev. fr. Rech. Opérat.*, n° 8, 1968.

[17] *Ordres totaux finis, travaux du séminaire sur les ordres totaux finis*, Aix-en-Provence/Paris, Mouton/Gauthier-Villars, 1971.