*R. franç. sociol.* XXXVI, 1995, 743-757

**Linton C. FREEMAN**

# Un modèle de la structure des interactions dans les groupes

RÉSUMÉ

En 1977, Winship montre que lorsque les proximités sociales reliant un ensemble d'individus satisfont les propriétés d'une ultramétrique, la structure sociale qui donne lieu à ces proximités est une hiérarchie ascendante dans laquelle un ensemble d'acteurs sont rassemblés par agrégations successives. Dans la mesure où cette forme structurale correspond à l'une des conceptions sociologiques minimales du groupe social, il est important de disposer d'un modèle permettant d'établir dans quelle mesure les proximités sociales observées au cours d'enquêtes empiriques remplissent les conditions d'une ultramétrique. Cet article propose un modèle stochastique qui utilise des données sur sept petits ensembles sociaux différents pour montrer que la structure des interactions s'y rapproche de la forme ultramétrique.

## Modéliser la structure des interactions dans les groupes

La notion de « groupe », telle qu'elle est utilisée par les sociologues, suppose un certain nombre de caractéristiques. La théorie sociologique voit dans les groupes des ensembles d'individus relativement petits, interagissant entre eux plus qu'avec des individus extérieurs, développant des relations affectives, se reconnaissant dans une identité commune et établissant des normes spécifiques. Mais la plupart des sociologues reconnaissent que l'interaction est ici essentielle au sens où les autres caractéristiques définissant le groupe ne peuvent émerger à moins que les interactions n'aient lieu (Bales, 1950, p. 33 ; Homans, 1950, p. 84 ; Flament, 1963, p. 47 ; Olmstead et Hare, 1978, pp. 10-11 ; Shaw, 1981, p. 8 ; Freeman, 1992). C'est pourquoi on a assisté, depuis les années 1940, à des efforts considérables visant à développer des modèles capables de spécifier les types de structures d'interaction associées aux collectivités identifiées sous le nom de groupes (1). Cet article s'inscrit dans cette ligne

(1) Les tentatives les plus anciennes de développement de modèles de la structure des interactions dans un groupe sont dues à Festinger (1949), Luce et Perry (1949) et Luce (1950). Mais ce n'est qu'au cours des années 1960 qu'un effort collectif systéma-

de recherche. J'y introduis une procédure permettant d'examiner dans quelle mesure un modèle structural particulier, fondé sur une classification hiérarchique, correspond bien à la structure des interactions observées.

On a aussi soutenu que les interactions structurées au sens défini ici émergent du fait de pressions sociales et psychologiques encourageant les acteurs à se comporter de manière à ce que les amis de leurs amis soient aussi des amis (Flament, 1963; Davis, 1967; Lorrain et White, 1971). Prenons, par exemple, un ensemble fini non vide contenant n acteurs sociaux, $A = \{x, y, ...\}$, ainsi qu'une relation binaire F dans $A \times A$. Une paire d'acteurs $(x, y) \in F$ si et seulement s'ils sont amis. Par définition, ce type de relation amicale est supposé symétrique de sorte que $(x, y) \in F \Leftrightarrow (y, x) \in F$. De plus, l'intuition que les amis d'amis sont des amis suggère que F doit être transitive, $(x, y) \in F$ & $(y, z) \in F \Rightarrow (x, z) \in F$. Dans l'un des premiers modèles de la structure de groupe, Davis (1967) a prouvé que, sur la base de ces conditions, la relation F distribue les acteurs de A en classes d'équivalence, ou groupes. Au sein de chaque groupe, tous sont amis les uns avec les autres et aucun membre d'un groupe n'est ami avec un membre d'un autre groupe.

Le problème avec cette formulation est bien sûr qu'il est difficile de concevoir des interactions amicales en termes de tout ou rien (Festinger, 1949; Levi-Strauss, 1963; Hubbell, 1965; Doreian, 1969; Lorrain et White, 1971; Alba, 1973; Peay, 1974, 1976, 1980; Seidman, 1983a, 1983b; Marsden et Laumann, 1984; Yan, 1988). Intuitivement, certaines paires d'acteurs sont proches et interagissent beaucoup, alors que d'autres sont moins proches et interagissent moins. Ainsi, comme on le voit généralement, ce type d'interaction amicale peut être quantifié. On peut concevoir que chaque paire d'acteurs peut être reliée par plus ou moins d'amitié. Nous avons donc besoin d'un modèle qui reste sensible à la quantité d'interactions caractérisant la relation entre individus.

Les sociologues ont beaucoup spéculé sur les conséquences, pour la structure du groupe, des variations individuelles dans la quantité d'interactions. Davis, Gardner et Gardner (1941, p. 150), par exemple, décrivent ces différences en termes de degré auquel les individus sont membres de leur groupe : « Ceux qui participent ensemble le plus souvent et aux affaires les plus intimes sont appelés des membres du noyau du groupe ; ceux qui

*(suite de la note 1)*
tique a été consenti dans cette direction. Les contributions à cet effort collectif sont dues à Flament (1963), Hubbell (1965), Davis (1967), Doreian (1969), Lorrain et White (1971), Alba (1973), Peay (1974, 1975, 1976, 1980), Winship (1977), Seidman et Foster (1978), Mokken (1979), Batchelder et Lefebvre (1982), King et Nakornchai (1982), Seidman (1983a, 1983b), Sailer et Gaulin (1984), Yan (1988), Arabie et Carroll (1989), Borgatti, Everett et Shirey (1990) et Freeman (1992, 1993). Le but de ces travaux a été de spécifier certaines propriétés structurales des ensembles d'individus que les sociologues appellent des groupes. Bien que cela ne soit pas courant en sociologie, cette ligne de recherche a donné lieu à un processus cumulatif (Hummon et Carley, 1993).

participent à certaines occasions avec ces membres du noyau, mais jamais en tant que groupe eux-mêmes, sont appelés des membres « primaires » ; alors que les individus aux marges, qui ne participent pas fréquemment, constituent des membres « secondaires ».

Lorsqu'on pense l'appartenance en termes de noyau et de périphérie, un modèle pertinent du groupe doit le représenter au moyen d'une sorte de structure hiérarchique arborescente. C'est précisément ce que fait, par exemple, le modèle introduit par Winship (1977). Ce modèle est fondé sur une application p de chaque paire d'acteurs dans A × A vers un sous-ensemble de nombres réels. Pour toute paire d'acteurs x et y, la valeur $p_{xy}$ est un indice de proximité sociale. D'après le modèle de Winship, ces proximités sociales doivent être ultramétriques, à savoir que pour tout x, y et z dans A, trois conditions doivent être remplies (Benzécri, 1973 ; Degenne et Forsé, 1994) :

(i) $p_{xx} > p_{xy}$

(ii) $p_{xy} > p_{yx}$

(iii) $p_{xy} \geq \min(p_{xz}, p_{yz})$.

D'après la condition (i), tout acteur doit être plus proche de lui-même ou d'elle-même que d'autrui. La condition (ii) spécifie que ces proximités sociales doivent être symétriques. La troisième exige que dans tout triplet d'acteurs, aucune paire ne peut être moins proche que le minimum des deux autres paires.

Winship a prouvé que lorsque ces conditions sont remplies, la structure qui en résulte est une hiérarchie arborescente, dans laquelle des sous-ensembles d'acteurs ne se chevauchent que par agrégations successives. De sorte que ce modèle conserve les éléments-clés de la conception sociologique traditionnelle des structures d'interaction.

L'une des implications du modèle ultramétrique permet d'appliquer un simple test de son ajustement à des proximités sociales observées. Considérons l'ensemble de tous les triplets non ordonnés d'acteurs (x, y, z) dans A. Chacun de ces triplets engendre trois paires non ordonnées d'acteurs (x, y), (y, z) et (x, z). Chacune de ces paires est associée avec une proximité sociale $p_{xy}$, $p_{yz}$ et $p_{xz}$ reliant ces acteurs. Assignons les labels a, b et c à ces trois proximités de manière à ce que a ≥ b ≥ c. La condition (iii) ci-dessus implique que, pour tous ces triplets, (a, b, c), c = b.

Cette dernière exigence est très stricte. On n'est donc pas surpris de voir que lorsque le modèle de Winship est confronté à des données observées de proximités sociales, l'ajustement n'est pas concluant (Freeman, 1992). En effet, comme Winship (1977) l'a lui-même prévu, la valeur observée de c est dans presque tous les cas inférieure à celle de b.

Le modèle de Winship peut être compris comme une sorte de structure idéalisée. Avec des données réelles, on ne s'attendrait pas à ce que les valeurs observées de c soient exactement égales à celles de b. Mais si le raisonnement sous-jacent au modèle est correct, on devrait néanmoins s'attendre à ce que les valeurs de c se rapprochent de celles de b. Dans la section suivante, j'introduis une procédure qui nous permet de déterminer dans quelle mesure ce rapprochement a lieu.

## Un modèle stochastiquement ultramétrique de la proximité sociale

Comme le modèle de Winship, le nôtre commence avec une matrice empirique de proximités sociales. On s'intéresse ici à la question de savoir dans quelle mesure ces dernières se rapprochent d'une forme ultramétrique. Lorsque les niveaux observés de c sont beaucoup plus petits que les niveaux correspondants de b, notre matrice empirique n'a clairement pas une forme ultramétrique. Mais dans la mesure où les niveaux observés de c sont presque aussi grands que les niveaux correspondants de b, notre matrice empirique se conforme davantage aux contraintes d'une structure ultramétrique.

Pour tout niveau observé de b, nous pouvons calculer le niveau moyen de c. Les proximités sociales observées se rapprochent alors de la forme ultramétrique dans la mesure où, pour tout niveau donné de b, le niveau moyen de c est beaucoup plus grand que ce à quoi l'on pourrait s'attendre au hasard. Pour préciser cette idée, nous avons besoin d'une procédure pour déterminer les valeurs au hasard de c.

Nous supposerons que notre distribution observée de $p_{xy}$ a été enregistrée sans erreurs. Etant donnée cette distribution, nous pouvons nous demander si ces $p_{xy}$ sont assignés à des paires (x, y) particulières de manière à ce que, pour un niveau donné quelconque de b, le niveau moyen de c soit plus grand qu'il ne le serait si les $p_{xy}$ étaient assignés à des paires (x, y) au hasard. Dans la mesure où ces niveaux moyens observés de c sont significativement plus grands que ceux auxquels on s'attendrait dans une distribution au hasard, ils tendent vers une forme ultramétrique. En d'autres termes, ils sont « stochastiquement ultramétriques ».

Considérons des données d'observation de proximités sociales reliant n acteurs. Dans de telles données on a

$$\frac{n(n-1)(n-2)}{3!} = \frac{n^3 - 3n^2 + 2n}{6}$$

triplets non ordonnés d'acteurs. On l'a vu, chaque triplet d'acteurs engendre trois proximités sociales. On se centre dans les deux modèles décrits plus haut sur ces triplets de proximités – qui relient des triplets d'acteurs.

Parmi les n acteurs, cependant, il y a

$$m = \frac{n(n-1)}{2!} = \frac{n^2 - n}{2}$$

paires non ordonnées. Et chacune de ces paires est reliée par une proximité sociale observée. De sorte que les données contiennent en fait

$$\frac{m^3 - 3m^2 + 2m}{6}$$

triplets non ordonnés de proximités reliant des paires de n acteurs.

Parce qu'il se centre seulement sur les proximités reliant des triplets d'*acteurs*, le modèle ultramétrique n'avance de prédictions que pour un petit sous-ensemble de tous les triplets de proximités. Le modèle incorpore l'idée que les proximités reliant des triplets d'acteurs sont sujettes aux contraintes structurales décrites plus haut. En effet, si les pressions sociales ou psychologiques conduisent au type d'équilibre où les amis des amis sont poussés à devenir amis, alors il doit résulter de ces pressions une situation dans laquelle c s'approche de b. Dans ces conditions, *la vaste majorité de triplets de proximité comprennent plus de trois acteurs et ne sont donc pas soumis à ce type de contrainte.* Le triplet de proximité $(p_{xy}, p_{xz}, p_{xw})$, par exemple, engage les quatre acteurs x, y, z et w. Le modèle n'impose aucune restriction sur les valeurs relatives de ces proximités.

Le fait que la grande majorité des triplets de proximité soient – en tout cas d'après la théorie – non contraints, suggère un moyen pratique de spécifier une hypothèse nulle. Appelons M(c|b) le niveau moyen de c étant donné b, parmi les trois proximités reliant les triplets d'acteurs dans les données considérées. Appelons μ(c|b) le niveau moyen correspondant parmi tous les triplets de proximités dans les données considérées. Ce calcul est équivalent au calcul de M(c|b) pour tous les recodages des $p_{xy}$ en termes de paires (x, y) auxquelles elles sont assignées.

Dans la mesure où les amis d'amis n'ont pas de relation particulièrement proche, M(c|b) devrait être approximativement égal à μ(c|b) et les proximités reliant des triplets d'acteurs ne devraient pas être différentes. Ainsi, dans la moitié des cas environ M(c|b) devrait être plus grand que μ(c|b) et dans l'autre moitié plus petit.

Cependant, dans la mesure où des amis d'amis sont reliés de manière plus proche que ce à quoi l'on pourrait s'attendre d'après l'hypothèse nulle, M(c|b) devrait être systématiquement plus grand que μ(c|b). Si ceci était le cas, il serait possible de conclure que les données observées tendent en effet vers une forme ultramétrique.

Un indice naturel du degré de différence entre M(c|b) et μ(c|b) est donné par la simple différence d = M(c|b) – μ(c|b). Cette différence est positive si $M(c|b)$ est plus grand que $\mu(c|b)$, elle est nulle s'ils sont égaux et négative dans le reste des cas.

Nous nous intéressons ici seulement aux différences positives – où, pour un niveau donné de b, le niveau moyen de c, $M(c|b)$, se rapproche du niveau de b. Cependant, le degré auquel $M(c|b)$ se rapproche de b dépend de la distance totale entre $\mu(c|b)$ et b. Si $\mu(c|b)$ et b sont proches, d tombe dans leur intervalle étroit; ses variations sont nécessairement limitées. Mais si $\mu(c|b)$ et b sont éloignés, d peut beaucoup varier. Il est donc utile de normaliser d et de l'exprimer comme une proportion de la valeur positive maximum qu'il peut prendre :

$$d' = \frac{M(c|b) - \mu(c|b)}{b - \mu(c|b)}$$

C'est cette mesure normalisée qui est utilisée dans l'analyse présentée ci-dessous.

On peut examiner la distribution des d' en utilisant des méthodes statistiques ordinaires. On définit une population composée de tous les triplets de proximité. On prend un échantillon composé seulement de triplets de proximité représentant des triplets d'acteurs. Pour tout niveau donné de b, on peut calculer la moyenne $\mu(c|b)$ et l'écart-type $\sigma(c|b)$ pour la population entière. Pour un même niveau de b, on peut déterminer le nombre $n(c|b)$ de triplets fondés sur des acteurs et la moyenne $M(c|b)$ pour l'échantillon de triplets fondés sur des acteurs. Pour ce niveau de b, donc, nous pouvons utiliser des statistiques fondées sur une distribution normale afin de déterminer la vraisemblance qu'un échantillon de taille $n(c|b)$ avec une moyenne $M(c|b)$ émerge d'une population caractérisée par une moyenne $\mu(c|b)$ et un écart-type $\sigma(c|b)$.

## Ajustement du modèle aux données

Ce modèle a été testé sur sept fichiers de données, devenus classiques, que j'ai utilisés dans une étude antérieure (Freeman, 1992). Ces fichiers sont décrits au bas des Tableaux I à VII; ils comprennent des données sur la participation d'un groupe de femmes à des activités sociales informelles dans une communauté du Sud des Etats-Unis, *Sourthern Women data* (Davis, Gardner et Gardner, 1941) (2); des donnés fondées sur l'observation des interactions entre usagers d'une plage publique en Californie, *Beach data* (Freeman, Freeman et Michaelson, 1988); des données d'ob-

(2) *Note du traducteur.* Nous avons conservé les noms anglais des fichiers de données dans la mesure où ce sont les dénominations sous lesquelles on les trouve dans les logiciels du type UCINET et STRUCTURE.

servation de « qui déjeune avec qui » parmi les enseignants d'une université californienne, *Lunch data,* (Freeman, 1987) ; des données d'observation de « qui a des activités sociales avec qui » parmi les membres d'un club universitaire de karaté, *Karate data* (Zachary, 1977) ; des données d'observation de la fréquence d'interactions entre les membres d'un club d'étudiant, *Frat data,* d'un bureau, *Office data,* et d'un centre de recherche technologique, *Tech data,* (Bernard et Killworth, 1977 ; Bernard, Killworth et Sailer, 1980).

Pour chaque fichier de données, on a déterminé le niveau moyen observé de c pour chaque niveau de b. Puis on a calculé dans chaque cas la distribution de c pour chaque niveau observé de b sous hypothèse nulle. Ceci a permis de calculer d' et le z-score associé (3). Ces résultats sont présentés dans les *Tableaux I* à *VII.*

Les scores d' pour les sept fichiers indiquent que les niveaux observés c sont presque toujours plus grands que ce à quoi l'on pourrait s'attendre sous hypothèse nulle. Sur les 78 valeurs de d' calculées, seulement trois sont négatives. Dans la mesure où, sous hypothèse nulle, 39 d'entre elles auraient dû être négatives, la thèse d'une tendance vers une structure ultramétrique semble confortée.

TABLEAU I. – ***Les données* Lunch.**

| Observé b | Observé c | Attendu c | d' | z |
|---|---|---|---|---|
| 9 | 9,000 | 0,517 | 1,000 | 6,770 |
| 8 | 6,000 | 0,489 | 0,734 | 6,680 |
| 7 | 5,000 | 0,472 | 0,694 | 4,050 |
| 6 | 4,800 | 0,449 | 0,784 | 9,160 |
| 5 | 3,120 | 0,380 | 0,593 | 15,130 |
| 4 | 2,391 | 0,293 | 0,566 | 14,580 |
| 3 | 1,857 | 0,231 | 0,587 | 11,480 |
| 2 | 0,974 | 0,179 | 0,437 | 11,890 |
| 1 | 0,201 | 0,072 | 0,139 | 9,240 |

Ces données ont été récoltées par Sue Freeman (1987) qui a observé « qui déjeunait avec qui » dans le corps professoral de la School of Social Sciences de l'Université de Californie à Irvine. Cette Faculté n'est pas formellement organisée en départements et les bureaux des professeurs sont dispersés dans l'immeuble sans distinction de discipline. Certaines disciplines sont plus fortement organisées que d'autres, mais la plupart des enseignants semblent interagir avec des gens de disciplines différentes. Il y avait trois restaurants sur le campus au moment de la récolte des données. Tous trois ont été surveillés au moment du déjeuner pendant trente jours. Pendant cette période, 28 membres du corps enseignant ont été observés. Les données utilisées ici décrivent la fréquence de leurs déjeuners en commun.

(3) z est calculé de manière standard. Le niveau de z n'a pas été augmenté par une correction prenant en compte le caractère fini de la population.

TABLEAU II. *Les données* **Beach**.

| Observé b | Observé c | Attendu c | d' | z |
|---|---|---|---|---|
| 28 | 21,000 | 1,041 | 0,740 | 8,100 |
| 20 | 16,500 | 0,998 | 0,816 | 9,570 |
| 17 | 16,000 | 0,961 | 0,938 | 6,980 |
| 15 | 8,000 | 0,920 | 0,503 | 6,080 |
| 14 | 6,333 | 0,898 | 0,415 | 4,840 |
| 13 | 5,692 | 0,863 | 0,398 | 9,500 |
| 12 | 8,000 | 0,831 | 0,642 | 8,280 |
| 11 | 4,500 | 0,814 | 0,362 | 6,200 |
| 10 | 4,056 | 0,776 | 0,356 | 8,830 |
| 9 | 3,333 | 0,737 | 0,314 | 5,310 |
| 8 | 2,364 | 0,711 | 0,227 | 3,920 |
| 7 | 3,118 | 0,677 | 0,386 | 7,620 |
| 6 | 2,628 | 0,614 | 0,374 | 11,130 |
| 5 | 1,436 | 0,529 | 0,203 | 8,720 |
| 4 | 1,224 | 0,447 | 0,219 | 9,560 |
| 3 | 0,973 | 0,370 | 0,229 | 11,710 |
| 2 | 0,493 | 0,249 | 0,139 | 12,310 |
| 1 | 0,120 | 0,102 | 0,020 | 2,480 |

Ces données ont été récoltées en 1986 par Freeman, Freeman et Michaelson (1988) parmi les membres d'une communauté de véliplanchistes sur une plage de Californie du Sud. La pratique de la planche à voile y a commencé en 1977. Avec le temps, des nouveaux venus *ont rejoint* la communauté jusqu'à ce qu'elle ait à peu près 30 à 40 membres en 1984. Les données sur « qui interagit avec qui » ont été récoltées par observation systématique sur la plage pendant une période de 31 jours. Pendant au moins deux demi-heures par jour, on a enregistré le temps pendant lequel chaque véliplanchiste interagissait avec chaque autre véliplanchiste. Ces observations systématiques de fréquences d'interaction entre 34 véliplanchistes sont utilisées dans ma présente analyse.

De plus, les différences entre niveaux observés et attendus de c sont souvent très grandes. Le niveau moyen de d' pour les données *Lunch* est de 0,615. Ceci indique que, à tous les niveaux de b, le niveau observé de c est descendu à 61,5 % de la distance qui sépare le niveau attendu (sous l'hypothèse nulle) de la limite supérieure. Pour les données *Beach*, le chiffre correspondant est de 0,406, pour les données *Frat* de 0,367, pour les données *Southern Women* de 0,547, pour les données *Karate* de 0,128, pour les données *Office* de 0,177 et pour les données *Tech* de 0,230.

De plus, les Tableaux I à VII montrent que les valeurs d' croissent, pour la plupart, de façon monotone avec les niveaux de b. Seules les données *Office* montrent une tendance générale des valeurs moyennes de c à décroître à mesure que les valeurs de b croissent.

Les tableaux montrent également les z-scores associés à chaque niveau de b. Un z-score de 2,33 ou plus grand est significatif avec une probabilité de 0,01 dans un test unilatéral. Tous les z-scores, à tous les niveaux de b, dépassent ce niveau dans les données *Lunch*. Tout les z-scores sauf un (celui au niveau le plus bas de b) sont significatifs dans les données *Beach*

TABLEAU III. *Les données* **Frat**.

| Observé b | Observé c | Attendu c | d' | z |
|---|---|---|---|---|
| 40 | 30,000 | 1,873 | 0,738 | 9,220 |
| 23 | 19,000 | 1,815 | 0,811 | 8,650 |
| 20 | 18,000 | 1,798 | 0,890 | 5,880 |
| 19 | 11,714 | 1,771 | 0,577 | 9,850 |
| 18 | 14,000 | 1,751 | 0,754 | 4,690 |
| 17 | 10,000 | 1,742 | 0,541 | 6,400 |
| 16 | 7,400 | 1,729 | 0,397 | 4,980 |
| 15 | 6,765 | 1,704 | 0,381 | 8,410 |
| 14 | 5,839 | 1,661 | 0,339 | 9,800 |
| 13 | 1,667 | 1,632 | 0,003 | 0,030 |
| 12 | 5,527 | 1,595 | 0,345 | 13,110 |
| 11 | 5,212 | 1,541 | 0,388 | 12,580 |
| 10 | 3,778 | 1,485 | 0,269 | 11,450 |
| 9 | 3,824 | 1,415 | 0,318 | 14,160 |
| 8 | 3,291 | 1,346 | 0,292 | 12,700 |
| 7 | 2,482 | 1,276 | 0,211 | 11,890 |
| 6 | 2,355 | 1,176 | 0,244 | 16,520 |
| 5 | 1,989 | 1,050 | 0,238 | 19,690 |
| 4 | 1,451 | 0,897 | 0,179 | 17,250 |
| 3 | 1,062 | 0,691 | 0,161 | 20,560 |
| 2 | 0,651 | 0,624 | 0,020 | 2,770 |
| 1 | 0,247 | 0,257 | −0,013 | −0,440 |

Les données *Frat*, *Office* et *Tech* ont toutes été récoltées par Bernard, Killworth et Sailer (Bernard et Killworth, 1977; Bernard, Killworth et Sailer, 1980). Ces études fournissent des données seulement sur les interactions; aucune d'entre elles ne comprend des descriptions ethnographiques des individus qui en font partie. Les données *Frat* ont été récoltées lorsque les enquêteurs ont «loué» l'immeuble d'une « fraternité ». Cinquante huit étudiants de cette fraternité ont accepté de participer à une expérience sociale. Toutes les quinze minutes, vingt et une heures par jour pendant cinq jours, un observateur discret traversait les lieux publics de l'immeuble de la fraternité pour noter le nombre de fois où chaque paire de membres de la fraternité était vue en conversation.

et les données *Southern Women*. Tous sauf ceux des deux plus bas niveaux le sont dans les données *Karate*. Tous sauf ceux des deux plus bas niveaux et un à un niveau moyen de b le sont dans les données *Frat*. Pour ces cinq fichiers de données, donc, 53 z-scores sur 60 sont significatifs. Sur les sept qui ne le sont pas, six n'apparaissent qu'aux deux niveaux les plus bas de b.

Les choses se passent un peu moins bien avec les données *Tech* et *Office*. On n'a qu'un seul z-score non significatif dans les données *Tech*, mais il apparaît au plus haut niveau de b. On a quatre z-scores non significatifs dans les données *Office ;* à nouveau, ils apparaissent aux plus hauts niveaux de b. De plus, les z-scores moyens sont de 5,6 pour les données *Tech* et de 3,4 pour les données *Office*. Ces niveaux sont beaucoup plus bas que ceux observés dans les quatre premiers tableaux.

TABLEAU IV. – *Les données* **Southern Women**.

| Observé b | Observé c | Attendu c | d' | z |
|---|---|---|---|---|
| 6 | 5,714 | 1,943 | 0,930 | 7,890 |
| 5 | 5,000 | 1,833 | 1,000 | 2,860 |
| 4 | 3,128 | 1,696 | 0,622 | 9,070 |
| 3 | 2,130 | 1,456 | 0,437 | 8,060 |
| 2 | 1,375 | 1,090 | 0,313 | 7,970 |
| 1 | 0,582 | 0,590 | – 0,020 | – 0,250 |

Une équipe de quatre ethnographes décida dans les années 1930 d'observer la vie sociale dans une ville du sud des Etats-Unis. Ils y ont récolté ensemble des données pendant deux ans, puis trois d'entre eux (Davis, Gardner et Gardner, 1941) ont produit un rapport détaillé sur leur recherche. Ils s'intéressaient principalement à la stratification sociale. Ils ont essayé de découvrir les manières par lesquelles les habitants se catégorisaient les uns les autres en termes de classe sociale. Ils ont montré qu'il existait parmi ces Américains une ségrégation à la fois de couleur et de classe sociale. Les classes étaient les unités sociales les plus vastes au sein desquelles les gens s'autorisaient à développer des relations personnelles proches. Dans chaque classe, on distinguait des groupes d'interaction intime qu'ils ont appelés des « cliques ». Pour révéler des structures de clique, ils ont utilisé des données d'observations, des listes d'invités et des comptes rendus de journaux. Le corpus de données analysé ici est fondé sur l'étude de 14 petits événements sociaux informels qui ont permis de mesurer les fréquences d'interaction entre 18 dames de cette ville.

TABLEAU V. – *Les données* **Karate**.

| Observé b | Observé c | Attendu c | d' | z |
|---|---|---|---|---|
| 5 | 1,857 | 0,350 | 0,324 | 4,100 |
| 4 | 0,968 | 0,282 | 0,185 | 4,590 |
| 3 | 0,550 | 0,167 | 0,135 | 6,680 |
| 2 | 0,055 | 0,055 | 0,000 | 0,010 |
| 1 | 0,000 | 0,005 | – 0,005 | – 0,610 |

Entre 1970 et 1973, Zachary (1977) a récolté des données sur les membres d'un club universitaire de karaté. Du fait de l'accent mis sur les arts martiaux, le club employait un instructeur à temps partiel. Au début de l'étude, un conflit commençait à émerger entre l'instructeur et le président du club. L'instructeur voulait une augmentation de salaire et l'autorisation de faire payer ses cours. Le président pensait qu'il lui appartenait de déterminer les salaires des employés et qu'il fallait diminuer les coûts. Peu à peu les membres du club se sont rangés d'un côté ou de l'autre des protagonistes. Certains voyaient dans l'instructeur leur mentor, alors que d'autres ne voyaient en lui qu'un employé qui cherchait à les faire payer encore plus. Zachary a récolté des données systématiques sur les interactions entre les 34 membres du club dans des contextes autres que les réunions du club. La théorie de Zachary est que les individus qui interagissent dans beaucoup de contextes différents partagent davantage d'information et sont socialement plus proches que ceux qui interagissent dans peu ou pas de lieux différents. Il a donc enregistré le nombre de contextes différents dans lesquels il observait des interactions entre une paire d'individus. Ce sont ces données qui sont analysées ici.

TABLEAU VI. – ***Les données* Office**.

| Observé b | Observé c | Attendu c | d' | z |
|---|---|---|---|---|
| 11 | 4,000 | 0,679 | 0,322 | 2,290 |
| 10 | 8,000 | 0,661 | 0,786 | 7,470 |
| 9 | 0,667 | 0,639 | 0,003 | 0,040 |
| 8 | 1,500 | 0,598 | 0,122 | 2,120 |
| 7 | 0,571 | 0,550 | 0,004 | 0,050 |
| 6 | 1,389 | 0,512 | 0,195 | 3,770 |
| 5 | 0,918 | 0,452 | 0,131 | 4,570 |
| 4 | 0,674 | 0,348 | 0,123 | 6,300 |
| 3 | 0,315 | 0,229 | 0,049 | 3,420 |
| 2 | 0,114 | 0,086 | 0,031 | 3,960 |
| 1 | | | | |

Les données *Office* ont été récoltées dans un petit bureau de recherche en Virginie de l'Ouest. A nouveau, un observateur empruntait un itinéraire fixe à travers le bureau toutes les quinze minutes pendant deux périodes de deux jours. Le résultat est une matrice du nombre de conversations observées entre toutes les paires des quarante individus travaillant dans ce bureau.

TABLEAU VII. – ***Les données* Tech**.

| Observé b | Observé c | Attendu c | d' | z |
|---|---|---|---|---|
| 8 | 2,500 | 0,685 | 0,248 | 1,810 |
| 7 | 3,444 | 0,635 | 0,441 | 6,470 |
| 6 | 2,182 | 0,562 | 0,298 | 6,670 |
| 5 | 2,000 | 0,491 | 0,335 | 6,880 |
| 4 | 1,148 | 0,414 | 0,205 | 6,570 |
| 3 | 0,775 | 0,320 | 0,170 | 6,370 |
| 2 | 0,368 | 0,231 | 0,077 | 3,510 |
| 1 | 0,155 | 0,094 | 0,067 | 6,660 |

Les données *Tech* ont été récoltées dans un institut de technologie de l'Université de Virginie de l'Ouest. De nouveau, l'observateur circulait dans les lieux publics de l'institut, toutes les demi-heures pendant une semaine. Le résultat est une matrice du nombre d'interactions observées entre chaque paire d'individus appartenant à l'institut.

## Discussion des résultats

On peut faire plusieurs observations sur les résultats de cette analyse. La première, et peut-être la plus intéressante, est que ces résultats indiquent que Davis, Gardner et Gardner (1941) et Homans (1950) avaient presqu'entièrement raison en affirmant que les interactions amicales ont une structure hiérarchique ou arborescente. Nos résultats montrent que les proximités sociales tendent vers le type de forme ultramétique qu'implique ce raisonnement.

Deuxièmement, ils confirment une autre croyance traditionnelle en sociologie. Les sociologues pensent généralement que la tendance des amis d'amis à être des amis dépend de la force des relations entre acteurs (Rapoport, 1954 ; Granovetter, 1973 ; Freeman, 1992). De ce point de vue, on pourrait s'attendre à ce que deux individus x et z aient plus de chances d'être proches lorsqu'ils sont tous deux très proches d'un troisième individu y. Si, cependant, leurs liens avec y sont faibles ou modérés, le fait qu'ils soient tous deux reliés avec y aura peu ou pas d'impact sur la force du lien les reliant l'un à l'autre. La tendance générale des niveaux moyens de c à croître de manière monotone avec de hauts niveaux de b confirme certainement cette croyance.

Il faut noter cependant que, à l'exception des plus bas niveaux de b, la différence entre c moyen et son niveau attendu est statistiquement significative. En effet, les résultats des Tableaux I à VII indiquent que les liens modérés ou même plutôt faibles ont un effet, petit mais significatif, sur la tendance des niveaux moyens de c à s'éloigner des niveaux attendus. Ce n'est qu'aux niveaux les plus faibles de b que les z-scores deviennent non significatifs. Ceci suggère que la tendance des triplets d'acteurs à équilibrer leurs proximités s'étend à presque tout l'éventail des forces du lien. Toutes les relations, sauf les plus superficielles, semblent tendre dans la direction de l'équilibre. Il semblerait que non seulement les amis d'amis tendent à être amis, mais que les connaissances de connaissances tendent à être des connaissances.

Troisièmement, nos résultats confortent la thèse, présentée notamment par Marsden et Campbell (1984), suivant laquelle des contraintes institutionnelles ont un effet observable sur la structuration des relations. Ici, nous voyons que, bien qu'ils soient la plupart du temps stochastiquement ultramétriques, les niveaux de d' produits par les données *Office* et *Tech* sont en moyenne beaucoup plus bas que ceux produits par les données *Southern Women, Beach, Lunch* et *Frat.* Les données ethnographiques (Freeman, 1992) suggèrent que les deux premiers corpus de données ont été récoltés dans des contextes où des contraintes organisationnelles sont fortement présentes, contrairement aux quatre derniers corpus. Ces différences suggèrent que Marsden et Campbell ont raison lorsqu'ils affirment que ces contraintes peuvent modifier une structure relationnelle de manière importante. De plus, les niveaux relativement faibles de d' que l'on trouve dans les données *Karate* renvoient au fait qu'elles ont été récoltées pendant une période de conflit dans le club ; une bonne part de la structure observée pourrait donc refléter des préoccupations instrumentales plutôt que des interactions amicales.

Le modèle stochastique ultramétrique s'ajuste donc bien aux données dans une très grande majorité des cas. Dans 57 des 60 niveaux de b, dans les cinq corpus non soumis à des contraintes organisationnelles, et dans tous les 18 niveaux de b, dans les données soumises à des contraintes organisationnelles *(Tech* et *Office)*, les niveaux observés de c dépassent leurs niveaux attendus. Il est donc raisonnable de conclure que ce modèle stochastique s'ajuste relativement bien aux données de proximités sociales, en même temps qu'il nous en apprend beaucoup sur la structuration des relations humaines. Ces résultats suggèrent aussi que les procédures ultramétriques devraient être utilisées dans l'analyse des proximités sociales. Parce qu'elle impose une structure ultramétrique aux données, la classification hiérarchique apparaît comme une bonne méthode pour décrire la structure des ensembles cohésifs que nous avons appelés ici des groupes.

Pour conclure, la prudence s'impose : dans la mesure où je n'ai trouvé que sept petits corpus de données pertinentes du point de vue de la présente analyse, les résultats décrits doivent être considérés comme provisoires. Il nous faudra reproduire ces tests sur d'autres données de fréquence d'interaction avant de parvenir à des conclusions fermes sur la capacité des classifications hiérarchiques ascendantes à modéliser la structure de ces ensembles.

**Linton C. FREEMAN**
*School of Social Sciences, University of California*
*Irvine, CA 92717*

*Traduit par Emmanuel Lazega*

## RÉFÉRENCES BIBLIOGRAPHIQUES

**Alba R.D.**, 1973. – « A graph theoretic definition of a sociometric clique », *Journal of mathematical sociology*, 3, pp. 113-126.

**Arabie P., Carroll J.D.**, 1989. – « Conceptions of overlap in social structure », dans **L.C. Freeman, D.R. White** et **A.K. Romney** (eds), *Research methods in social network analysis*, Fairfax, VA, George Mason University Press.

**Bales R.F.**, 1950. – *Interaction process analysis : a method for the study of small groups*, Cambridge, MA, Addison-Wesley.

**Batchelder W.H., Lefebvre V.A.**, 1982. – « A mathematical analysis of a natural class of partitions on a graph », *Journal of mathematical sociology*, 26, pp. 124-148.

**Benzécri J.-P.**, 1973. – *L'analyse des données*, Paris, Dunod, vol. I : *La taxinomie*.

**Bernard H.R., Killworth P.D.**, 1977. – "Informant accuracy in social network data, II", *Human communication research*, 4, pp. 3-18.

**Bernard H.R., Killworth P.D., Sailer L.**, 1980. – « Informant accuracy in social network research, IV : a comparison of clique-level structure in behavioral and cognitive data », *Social networks*, 2, pp. 191-218.

**Borgatti S.P., Everett M.G., Shirey P.R.**, 1990. – « LS sets, lambda sets and other cohesive subsets », *Social networks*, 12, pp. 337-358.

**Davis A., Gardner B.B., Gardner M.R.**, 1941. – *Deep South*, Chicago, University of Chicago Press.

**Davis J.A.**, 1967. – « Clustering and structural balance in graphs », *Human relations*, 20, pp. 181-187.

**Degenne A., Forsé M.,** 1994. – *Les réseaux sociaux*, Paris, Armand Colin.

**Doreian P.,** 1969. – « A note on the detection of cliques in valued graphs », *Sociometry*, 32, pp. 237-242.

**Festinger L.,** 1949. – « The analysis of sociograms using matrix algebra », *Human relations*, 10, pp. 153-158.

**Flament C.,** 1963. – *Applications of graph theory to group structure*, Englewood Cliffs, NJ, Prentice-Hall, traduit en français sous le titre *Théorie des graphes et structure sociale*, Paris, Mouton-Gauthier Villars, 1965.

**Freeman L.C.,** 1984. – « Turning a profit from mathematics : the case of social networks », *Journal of mathematical sociology*, 10, pp. 343-360.

— 1992. – « On the sociological concept of 'group' : an empirical test of two models », *American journal of sociology*, 98, pp. 152-166.

— 1993. – « Finding groups with a simple genetic algorithm », *Journal of mathematical sociology*, 17, pp. 227-241.

**Freeman L.C., Freeman S.C., Michaelson A.G.,** 1988. – « On human social intelligence », *Journal of social and biological structures*, 11, pp. 415-425.

**Freeman S.C.,** 1987. – *Verbal response data as predictors of social structure*, PhD Dissertation, University of California, Irvine.

**Granovetter M.,** 1973. – « The strength of weak ties », *American journal of sociology*, 78, pp. 1360-1380.

**Homans G.C.,** 1950. – *The human group*, New York, Harcourt, Brace.

**Hubbell C.H.,** 1965. – « An input-output approach to clique identification », *Sociometry*, 28, pp. 377-399.

**Hummon N.P., Carley K.,** 1993. – « Social networks as normal science », *Social networks* 15, pp. 71-106.

**King J.R., Nakornchai V.,** 1982. – « Machine-component group formation in group technology : review and extension », *International journal of production research*, 20, pp. 117-133.

**Levi-Strauss C.,** 1963. – *Structural anthropology*, New York, Basic Books.

**Lorrain F.P., White H.C.,** 1971. – « Structural equivalence of individuals in social networks », *Journal of mathematical sociology*, 1, pp. 49-80.

**Luce R.D.,** 1950. – « Connectivity and generalized cliques in sociometric group structure », *Psychometrika*, 15, pp. 169-190.

**Luce R.D., Perry A.,** 1949. – « A method of matrix analysis of group structure », *Psychometrika*, 14, pp. 95-116.

**Marsden, P.V., Laumann E.O.,** 1984. – « Mathematical ideas in social structure analysis », *Journal of mathematical sociology*, 10, pp. 271-294.

**Marsden P.V., Campbell K.E.,** 1984. – « Measuring tie strength », *Social forces*, 63, pp. 482-501.

**Mokken R.J.,** 1979. – « Cliques, clubs and clans », *Quantity and quality*, 13, pp. 161-173.

**Olmstead M.S., Hare A.P.,** 1978. – *The small group*, New York, Random House.

**Peay E.R.,** 1974. – « Hierarchical clique structures », *Sociometry* 37, pp. 54-65.

— 1975. – « Nonmetric grouping : clusters and cliques », *Psychometrika*, 40, pp. 297-313.

— 1976. – « A note concerning the connectivity of social networks », *Journal of mathematical sociology*, 4, pp. 319-321.

— 1980. – « Connectedness in a general model for valued networks », *Social networks*, 2, pp. 385-410.

**Rapoport A.,** 1954. – « Spread of information through a population with socio-structural bias, III : Suggested experimental procedures », *Bulletin of mathematical biophysics*, 16, pp. 75-81.

**Sailer L.D., Gaulin S.J.C.,** 1984. – « Proximity, sociality and observation : the definition of social groups », *American anthropologist*, 86, pp. 91-98.

**Seidman S.B.,** 1983a. – « LS sets and cohesive subsets of graphs and hypergraphs », *Social networks*, 5, pp. 92-96.

1983b. – « Internal cohesion of LS sets in graphs », *Social networks*, 5, pp. 97-107.

**Seidman S.B., Foster B.L.,** 1978. – « A graph theoretic generalization of the clique concept », *Journal of mathematical sociology*, 6, pp. 139-154.

**Shaw M.E.,** 1981. – *Group dynamics : the psychology of small group behavior*, New York, McGraw Hill, 3e ed.

**Winship C.,** 1977. – « A distance model for sociometric structure », *Journal of mathematical sociology*, 5, pp. 21-39.

**Yan X.,** 1988. – « On fuzzy cliques in fuzzy networks », *Journal of mathematical sociology*, 13, pp. 359-389.

**Zachary W.,** 1977. – « An information flow model for conflict and fission in small groups », *Journal of anthropological research*, 13, pp. 452-473.