

PHILIPPE RICHARD

Étude des renchaînements d'alliance

Mathématiques et sciences humaines, tome 123 (1993), p. 5-35

http://www.numdam.org/item?id=MSH_1993__123__5_0

© Centre d'analyse et de mathématiques sociales de l'EHESS, 1993, tous droits réservés.

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

NUMDAM

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

ÉTUDE DES RENCHAÎNEMENTS D'ALLIANCE

Philippe RICHARD¹

RÉSUMÉ — *Repéré depuis longtemps dans les sociétés rurales françaises, le renchaînement d'alliance ou mariage dans la parenté affine a été étudié dans un petit nombre de communes. Nous disposons, maintenant, de deux communes de référence qui permettent d'analyser ce phénomène dans un cadre comparatif. La question du choix des meilleures mesures de ce phénomène est discutée, compte tenu, en particulier, de la pauvreté de l'information dont on dispose, dans certains cas, de son hétérogénéité presque toujours. Les distributions de Pareto-Levy se révèlent efficaces et permettent de dépasser les premières intuitions. La distribution du nombre de renchaînements par ancêtre renchaineur est bien estimée par une des distributions de Pareto et, ramenée au nombre de descendants de chaque ancêtre, elle se révèle être un bon indicateur. La capacité d'une population à renchaîner est mise en relation avec le nombre de renchaînements d'un type donné qui peut être évalué malgré les insuffisances de la connaissance généalogique sur la population. Ces mesures sont appliquées à deux communes qui, grâce à leurs différences, permettent de vérifier l'intérêt de ces calculs. Des propriétés structurales globales du réseau de renchaînements sont discutées, montrant l'intrication élevée de ce réseau. Enfin, diverses pistes de recherche sont proposées pour prolonger l'étude de ce phénomène et en accroître la compréhension.*

SUMMARY — The relinking of lines. An approach by the calculus

Relinking of lines (renchaînement) or marriages between affines belonging to lines within kindreds has been found since a long time in French rural societies but its study has concerned so far only a limited number of townships (communes). There are now two reference townships which enable us to analyse that phenomenon within a comparative structure. The question of selecting the best measurements applied to this phenomenon is discussed here, taking into account in some cases the scarcity of the information available and, in all cases its heterogeneity. The Pareto-Levy laws appears to be efficient here and enable us to go beyond our first intuitions. Owing to one of the Pareto distributions we have been able to make an estimation of the number of relinking lines attributed to each ancestor responsible of the relinking. Once brought back to the number of descendants it appears to be a good indicator. The capacity of a population to resort to relinking of lines is correlated with the number of relinking lines of a certain type with the deduction of which is possible in spite of inadequate genealogical datas concerning the population concerned. The measurements are applied to two townships which, being very different, enable us to verify the relevance of the system of calculation. Global structural properties of the relinking network are discussed, showing the high degree of intricacy of the pattern. In the end various proposals for further research are given in order to allow a better understanding of that phenomenon.

¹ Cet article s'inscrit dans une série de recherches, initiées par Martine Ségalen à Saint Jean Trolimon, en Bretagne. Ces recherches visent à la fois à analyser les comportements matrimoniaux en fonction de cette structure de mariage et à en étudier les propriétés mathématiques. S'il est clair que le méthodologue ne peut se substituer à l'ethnologue, la nature et la précision des questions qu'il pose et des réponses qu'il apporte dépendent de la somme de connaissances mathématiques qui ont pu être accumulées sur les phénomènes et qui permettent d'en interpréter la physique.

1. PRÉSENTATION DU PROBLÈME

Dans les sociétés complexes selon la définition de Claude Levi-Strauss le choix du conjoint n'est pas réalisé selon des règles prescriptives. Les mariages entre consanguins proches (petits enfants, arrière petits enfants) étant sévèrement contrôlés par l'église, une certaine partie de l'endogamie que l'on observe dans les sociétés rurales ne peut s'expliquer que par des mariages entre consanguins proches ou lointains, des alliés.

Les ethnologues ont repéré depuis longtemps dans les sociétés rurales françaises la figure du renchaînement d'alliance, ou mariage dans la parenté affine. Le poids de cette configuration matrimoniale et des pratiques sociales qui la sous-tendent ne peut être clairement mesuré que si l'on dispose d'un corpus généalogique qui prend en compte les ascendances des deux lignées, puisque nos sociétés sont cognatiques². Un tel corpus a été rassemblé une première fois et traité par un programme informatique sur une population de basse Bretagne³. L'importance du renchaînement d'alliance, comme pratique systématique matrimoniale, y apparaissait clairement, singularisant, au sein de la population, un sous-ensemble de lignées qui renchaînaient très fortement les alliances. Des explications sociologiques étaient fournies de ces comportements.

L'entreprise a été répétée sur la commune de Nanterre, pour laquelle un corpus portant sur 7 générations, du 18^{ème} au 19^{ème} siècle, a été relevé. Bien que, dans un cas comme dans l'autre, les résultats aient été interprétés en l'absence de toute comparaison entre ces deux communes, il est clair que la reconnaissance du renchaînement d'alliance ne peut se développer hors d'un cadre comparatif. Le simple bon sens nous confronte à cette question: "Y a-t-il plus de renchaînements d'alliance à Saint Jean Trolimon qu'à Nanterre"?

Par ailleurs, est-il possible de maintenir un questionnement sur les fonctions sociales de ce phénomène sans chercher à définir plus précisément ce qu'on mesure? C'est sous la pression de cette interrogation que la recherche présentée ici est menée. Nous ne pensons pas qu'il soit possible, dès maintenant, de déterminer précisément quelle (ou quelles) mesures sont les plus caractéristiques du renchaînement d'alliance, mais au moins, la comparaison si limitée soit-elle, de deux communes et, partant, de deux interprétations du phénomène, permet-elle de poser les questions les plus fécondes et d'apporter quelques éléments de réponse.

Rappelons d'abord ce qu'est un renchaînement d'alliance. Il y a renchaînement d'alliance lorsque le mariage entre deux individus non consanguins est suivi du mariage entre deux consanguins des époux. Cette structure de mariage met donc en jeu deux couples $A=B$ et $C=D$, les ancêtres respectifs de A, C et B, D que nous nommerons M et N et les quatre liens de parenté MA, MC et NB, ND⁴.

² Se dit d'une société dans laquelle la parenté est transmise indifféremment par le père ou par la mère, au contraire des sociétés agnatiques qui définissent la parenté par les pères seuls. On peut parler aussi de parenté ou de filiation indifférenciée.

³ La fastidieuse construction des généalogies a été réalisée par Martine Ségalen et contrôlée au fur et à mesure. Les réflexions qui suivent sont le fruit d'une collaboration étroite et extrêmement enrichissante. Qu'elle trouve ici l'hommage d'un méthodologue qui a longuement profité de cette association.

⁴ Seul ce type de renchaînement d'alliance est étudié ici, à l'exclusion de celui qui fait intervenir entre A et C ou B et D un lien d'affinité. L'exemple le plus simple de cette structure est représentée par un ensemble de trois couples liés par des liens de germanité.

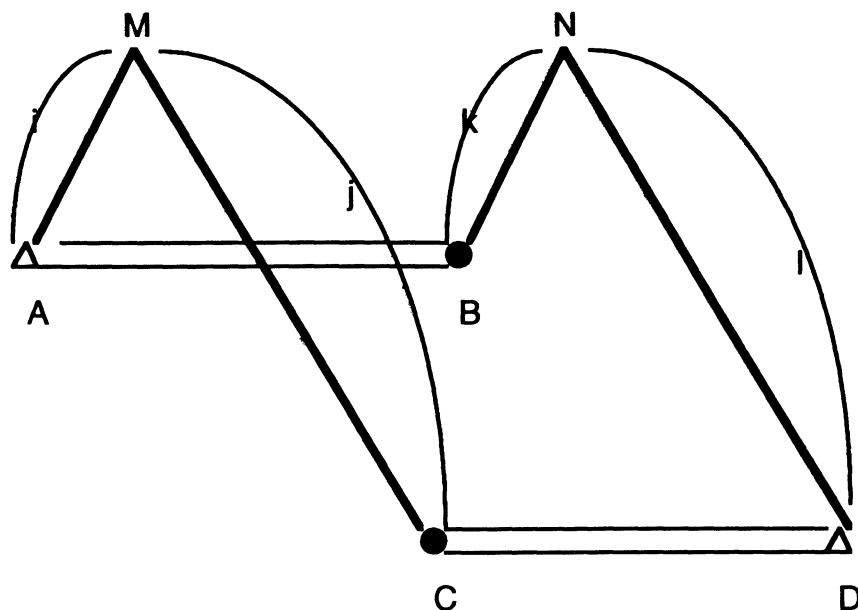


Figure 1. La structure élémentaire du renchaînement d'alliance

Une telle définition appelle quelques remarques.

1). Si A et B sont déjà consanguins par M, on montre facilement (figure 3) que D et C le sont aussi. En ce cas, le renchaînement d'alliance ressort plus à une étude sur la consanguinité dans le mariage.

2). Nous ne distinguons pas, pour le moment, les consanguins vrais des demi-consanguins. Deux descendants par le même père, mais de mères différentes (figure 2) sont traités comme si leurs mères étaient les mêmes.

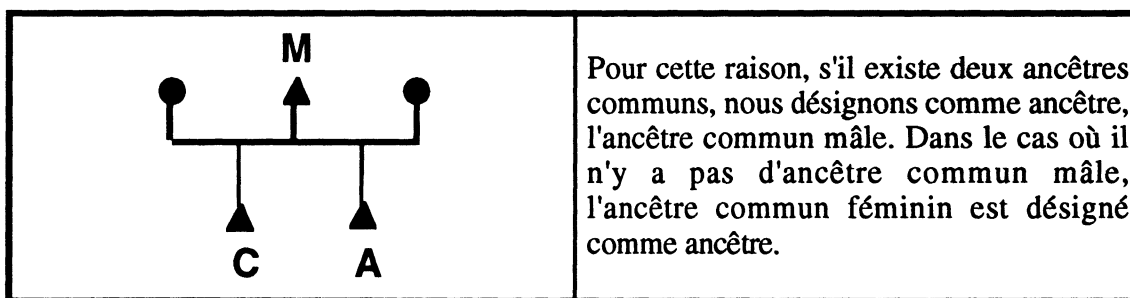


Figure 2. Le traitement des demi-consanguins

Constater dans une population la présence de mariages renchaînant les alliances est une chose. Déterminer si cette population est plus renchaîneuse que telle autre ou vérifier que les renchaînements de tel groupe socio-économique n'ont pas exactement la même structure que ceux d'un autre groupe implique que l'on puisse construire une ou plusieurs mesures de ce phénomène qui le représente de façon efficace et féconde.

Or, si dans la plupart des cas traités en démographie historique, la mesure des divers coefficients utilisés se fait avec toute l'information désirée, il est de la nature même de la figure du renchaînement d'alliance d'être incomplètement repérée dans une société donnée. En effet, pour extraire tous les p-renchaînements d'un ensemble d'individus, il faut disposer pour chaque individu de ses ascendants jusqu'à la p^{ème} génération. Ceci est rarement le cas - en particulier - pour les conjoints étrangers au groupe social représenté par l'ensemble. Il importe

donc de déterminer de quelle façon et dans quelle mesure le manque d'information peut-être dépassé pour disposer d'une connaissance exacte et comparative du phénomène.

Le but de cet article est donc de montrer les problèmes rencontrés dans la mesure et le calcul des renchaînements d'alliance et les solutions proposées. Par la comparaison des résultats obtenus pour deux populations différentes, nous montrerons quels coefficients construire, quelles fonctions analyser, pour avoir une bonne représentation de ce phénomène, la moins sensible possible aux manques d'information que l'on constate.

2. CONSTRUCTION ET CONTRÔLE DES GÉNÉALOGIES

2.1. *La construction des généalogies*

La constitution d'une généalogie s'étendant sur une ou deux dizaines de générations est une opération si fastidieuse et si onéreuse que l'on comprend que de telles entreprises soient rares. Dans les peuples occidentaux, les sources écrites sont toujours les archives, civiles ou paroissiales, de mariage, naissance et décès, qu'il faut mettre en relation les unes avec les autres en identifiant sans erreur chaque individu cité d'une source à l'autre ou d'un acte à l'autre. Depuis quelques années des programmes de reconstitution automatique de généalogies ont vu le jour⁵. Leur efficacité diminue quand le stock des noms et prénoms de la population est faible et quand les graphies désignant le même individu s'accroissent. D'autre part, l'efficacité de ces programmes repose sur la redondance des informations, à la fois différentes mais formant un tout cohérent. Leur utilisation nécessite une saisie d'informations abondantes. Le gain de temps qu'ils représentent est payé par un coût de saisie plus élevé. Aussi avons nous suivi une autre voie. Les seules informations dont nous avons besoin sont saisies et l'identification de l'individu est faite manuellement. Cette opération a été grandement facilitée par l'usage des logiciels de traitement de fichiers des microordinateurs (WORKS dans le cas présent). Ce qui explique que la généalogie nanterrienne a été plus aisée à construire que celle de Saint Jean Trolimon, faite sans de telles aides.

2.2. *Le contrôle des généalogies*

Les recherches sur les renchaînements d'alliance ne peuvent supporter d'erreurs dans la construction des généalogies. En effet, identifier par erreur un individu I comme étant un individu J, c'est transporter toute une partie de l'ascendance de J dans celle de I et multiplier les repèrages erronés de renchaînements d'alliance.

De la même façon, ne pas reconnaître que I et J sont le même individu, c'est se priver de tout un ensemble de renchaînements d'alliance. Il importe donc de pourchasser impitoyablement les erreurs d'identification, qu'elles soient dues à des confusions sur l'identité de deux individus ou à des erreurs factuelles de codes.

Pour parvenir à ce résultat, nous avons contrôlé deux types d'informations :

- 1) Tout code référence l'individu de façon unique. L'inverse est plus difficile à vérifier par suite de l'abondance des homonymes.
- 2) Le graphe des individus et de leurs ascendants est une arborescence. Pour tout individu I dans le graphe, nous vérifions qu'aucun ancêtre de I n'est I lui même.
- 3) Tout individu de sexe donné garde son sexe dans les différentes positions de parenté qu'il occupe (Mari/Père, Épouse/Mère).

⁵ Ces programmes utilisent la redondance de l'information pour désambiguïser les irrégularités ou les identités de dénomination. Ainsi, les informations sur les professions, les parrains et marraines, les surnoms alimentent les algorithmes de recherche qui sont d'autant plus efficaces que ces informations sont nombreuses. Leur emploi ne résolvant pas tous les problèmes, un contrôle manuel reste nécessaire.

4) Nous vérifions la cohérence des dates de naissance et des âges au mariage⁶, ainsi que la vraisemblance des intervalles entre les naissances des germains.

5) Tout individu intervenant dans la généalogie apparaît deux fois dans nos fichiers: d'abord en tant qu'Ego avec ses parents, conjoints et diverses dates. Les parents sont ensuite notés une deuxième fois en tant que conjoints mutuels. Le programme de contrôle vérifie la cohérence de ces informations redondantes.

Si l'on ajoute à ces contrôles systématiques, les vérifications que représente l'étude des exemples ou des données - tests de programme - on peut être confiant sur la consistance des généalogies ainsi étudiées.

2.3. La construction des générations

La recherche des renchaînement d'alliance ne peut se faire sans que l'ensemble des individus ne soit divisé en classes d'âge - les générations - postulées arbitrairement de 25 ans. Ce problème ne mériterait pas d'être évoqué si chaque individu avait une date de naissance connue avec précision. Comme ce n'est pas le cas, il importe de l'estimer. Nous avons appliqué les règles suivantes dans cet ordre.

- Deux germains sont de la même génération.
- La distance entre parents et enfants est d'une génération.
- Deux époux sont de la même génération.

De telles règles ne tiennent pas compte des grands intervalles aîné et cadet dans les fratries nombreuses. Elles ignorent les écarts d'âge, parfois élevés, entre conjoints⁷. Mais telles qu'elles sont, elles permettent de classer tout individu dans une génération et les biais qu'elles introduisent sont traités dans l'algorithme de recherche de renchaînements d'alliance.

2.4. La recherche des renchaînements d'alliance

2.4.1. La définition que nous avons donnée du renchaînement d'alliance est théorique. Pratiquement, nous ne rechercherons les consanguins susceptibles de participer à un renchaînement d'alliance que pour autant que l'ancêtre commun ne se situe pas à plus de quatre générations. Il peut être un arrière-arrière-grand-père au maximum. Cette restriction se justifie par l'hypothèse que nous formulons ainsi : les renchaînements d'alliance reposent sur des choix conscients. Compte tenu de la mémoire généalogique des populations françaises il est peu réaliste de considérer comme volontaire un choix qui implique de se reconnaître parents avec un cousin au 4ème degré.

Les structures de renchaînement d'alliance dans lesquelles se trouve un mariage consanguin sont reconnues mais non prises en compte dans l'ensemble des renchaînements d'alliance.

⁶ Les dates exprimées en calendrier révolutionnaire ont été converties en années. En effet, pour la définition des générations, l'année est une variable suffisamment précise. Certaines recherches, non développées dans cet article, utilisant un ordre sur les renchaînements, nécessiteraient la prise en compte de dates plus précises.

⁷ Il existe un groupe d'hommes qui se marient systématiquement après 65 ans. Ce sont généralement des militaires de profession. On peut, dans de tels cas, avoir des écarts entre conjoints qui atteignent 52 ans, soit deux générations.

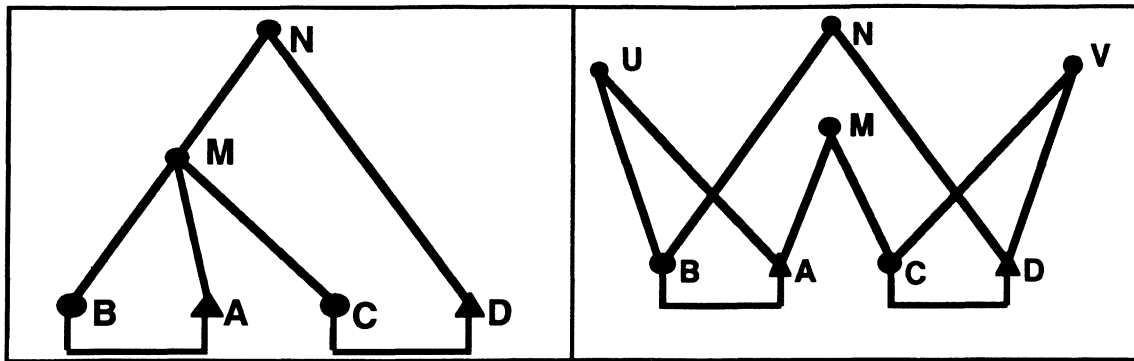


Figure 3 (a et b). Décomposition d'un renchaînement d'alliance par un mariage consanguin

2.4.2. Une structure de renchaînement d'alliance peut intégrer un ou deux mariages consanguins. On peut démontrer qu'une structure de renchaînement d'alliance dans laquelle il existe un mariage consanguin est une structure dans laquelle les deux mariages sont consanguins, si l'un des ancêtres appartient à la lignée de l'autre ou si les deux ancêtres sont confondus (figure 3a). L'autre cas, présenté dans la figure 3b, est encore plus rare : les deux couples sont consanguins indépendamment l'un de l'autre. Il s'agit, par exemple, de deux germains épousant deux germains, les deux paires de germains étant par ailleurs cousins. Nous traitons ces structures particulières comme des structures de consanguinité et non de renchaînement d'alliance. Ce choix ne fait que traduire une intuition sociologique selon laquelle le sentiment de consanguinité prend le pas sur celui de renchaînement d'alliance.

En conséquence nous repérons les structures de renchaînement d'alliance non consanguines et ne dépassant pas le 4^{ème} degré.

2.4.3. L'algorithme de recherche.

Disons quelques mots de l'algorithme de recherche car c'est lui qui rend effective la définition précédente.

1) Deux individus A et B sont dits consanguins s'il existe un individu M (un seul suffit) ancêtre de A et de B, à une distance généalogique égale ou inférieure à 4. Les demi-consanguins sont assimilés aux consanguins.

2) Pour tout A, de génération g

pour tout B conjoint de A

pour tout C de générations g-3 à g+3

pour tout D conjoint de C

il y a renchaînement d'alliance si :

A et C sont consanguins par M et B et D consanguins par N

Aucun des sommets A, B, C, D n'est un ancêtre de l'un de ces sommets

A et B sont non consanguins et C et D non consanguins.

2.4.4. Remarques :

L'ancêtre commun est toujours le conjoint mâle du couple ancêtre, sinon c'est l'ancêtre commun, unique, mâle ou femelle.

La recherche des consanguins se fait sur un intervalle de 7 générations pour tenir compte des décalages de générations pertinentes. En toute rigueur il faudrait chercher les renchaînements d'alliance sur 9 générations pour tenir compte des erreurs dans le calcul des générations. Le coût est très élevé pour un bénéfice minime.

On peut montrer que le nombre d'essais est inférieur à 4,4 millions pour une population de 5000 habitants, classée en 7 générations de 2×350 individus avec un nombre moyen de conjoints de 1,5⁸. Il importe de minimiser le temps de calcul dû à la comparaison de deux généalogies ascendantes. Ces généalogies ont été construites une fois pour toutes, pour chaque individu. La comparaison se réduit à la recherche d'un élément identique dans deux tableaux de 64 positions. L'algorithme est optimisé pour éviter les essais inutiles.

Un renchaînement d'alliance étant défini par la donnée de 6 sommets dont deux couples, 4 liens et un indice d'identité de sexe (sexe de C par rapport à celui de A), c'est l'ensemble de ces 13-uplets qui caractérise les renchaînements d'alliance d'une généalogie recueillie sur une population donnée. Ainsi Saint Jean Trolimon a fourni 2359 13-uplets de ce type, Nanterre en a fourni 1214. Toute l'information sur les renchaînements d'alliance tient dans ces ensembles.

2.5. La représentativité des généalogies par rapport à la population

Dire que l'ensemble de l'information sur les renchaînements d'alliance tient dans cette liste laisse ouvert le problème de la représentativité de la généalogie par rapport à la population. La comparaison de Nanterre et de Saint Jean Trolimon est, en ce sens, instructive. Les généalogies de Nanterre sont réalisées à partir de tous les actes de mariage inscrits dans les registres de la commune. Un couple dont l'un des conjoints provient d'une autre commune sera inscrit mais nous ne disposerons d'aucune information pour ce conjoint étranger et nous ne ferons aucune recherche dans les communes voisines.

Les généalogies de Saint Jean Trolimon ont été construites à partir d'un noyau de familles qui ont occupé leur ferme pendant au moins 3 générations. Pour chaque ferme, les généalogies des individus et de leur conjoint habitant ces fermes étaient reconstruites. Par contre les généalogies ascendantes seules étaient construites pour les enfants quittant la ferme.

Ainsi, nous connaissons bien les fratries à Nanterre, moins bien les ancêtres éloignés. A Saint Jean Trolimon nous disposons de généalogies profondes mais les enfants d'un couple sont souvent ignorés. Ces ignorances relatives doivent être prises en compte dans les raisonnements ultérieurs.

3. ÉTUDE DES RENCHAÎNEMENTS

3.1. Les facteurs analysés

L'étude des renchaînements peut porter sur les ancêtres par lesquels passent les renchaînements ou sur les couples renchaîneurs. Elle peut concerner l'importance de ces individus ou couples par rapport à l'ensemble. Mais elle peut porter aussi sur les liens de parenté entre consanguins. Enfin, elle peut s'intéresser aux relations structurelles que forment l'ensemble des couples reliés les uns aux autres par un renchaînement. Chacun de ces facteurs peut être mis en relation avec une propriété socio-économique des individus ou couples concernés afin de vérifier l'influence de ces propriétés sur le phénomène.

⁸ Si n est la taille moyenne d'une génération, g le nombre de générations, k l'intervalle de recherche et c le nombre moyen de conjoints par individu, le nombre d'essais à effectuer est majoré par :

$$g(kn)^2 / 4c^2$$

On néglige en effet, dans ce calcul simplifié, le fait que pour les générations 1 à 3 et $g-3$ à g l'intervalle de $2k+1$ générations ne peut être exploré entièrement et que les nombres de femmes et d'hommes ne sont pas identiques. La correction est importante pour g et k voisins. On néglige aussi que, pour chaque génération, les ascendants de l'individu ne sont pas essayés.

3.2. Étude de la distribution des ancêtres

Pour chaque renchaînement, deux ancêtres A et B sont mis en évidence. Certains apparaissent fréquemment, d'autres rarement. Nous appellerons poids brut (plus simplement dit poids quand il n'y a pas d'ambiguïté) d'un ancêtre A noté $p(A)$ le nombre de renchaînements dans lesquels il est inscrit, c'est à dire pour lequel il est un ancêtre. Soit $L(A)$ l'ensemble des descendants de A, fils et filles, petits enfants etc... et $l(A)$ le nombre d'individus composant cet ensemble. Le phénomène du renchaînement définit dans l'ensemble des individus le sous-ensemble de ceux qui interviennent au moins une fois dans un renchaînement en tant qu'ancêtre. On peut donc étudier la distribution de $p(x)$ pour tout x appartenant à ce sous-ensemble.

Avant de procéder à cette étude, remarquons :

- 1) le poids de A est corrélé positivement avec la taille $l(A)$ de $L(A)$. Il apparaît donc naturel, soit d'étudier la distribution de $p(x)/l(x)$, soit d'étudier $p(x)$ en fonction de $l(x)$.
- 2) $p(A)$ est une propriété de A, l'ancêtre, non de $L(A)$ sa descendance. Il se trouve des descendants de A, qui appartiennent à $L(A)$, qui sont eux-mêmes ancêtres de renchaînements. Ils sont non comptabilisés pour A. Seuls sont comptés pour A les renchaînements passant par A.
- 3) une première difficulté apparaît donc. L'étude doit-elle porter sur $p(A)$ ou sur $p^*(A)$, le nombre de renchaînements pour A et tout descendant de A ?

4. LA DISTRIBUTION DES POIDS BRUTS ET RÉDUITS

4.1 Étude de la distribution de $p(A)$

Le tracé de la distribution de la fréquence de $p(A)$ (nombre de sommets A tels que $p(A)$ a une valeur x) est manifestement de type hyperbolique - peu de sommets à poids élevé, de très nombreux sommets de poids très faible (figure 4)

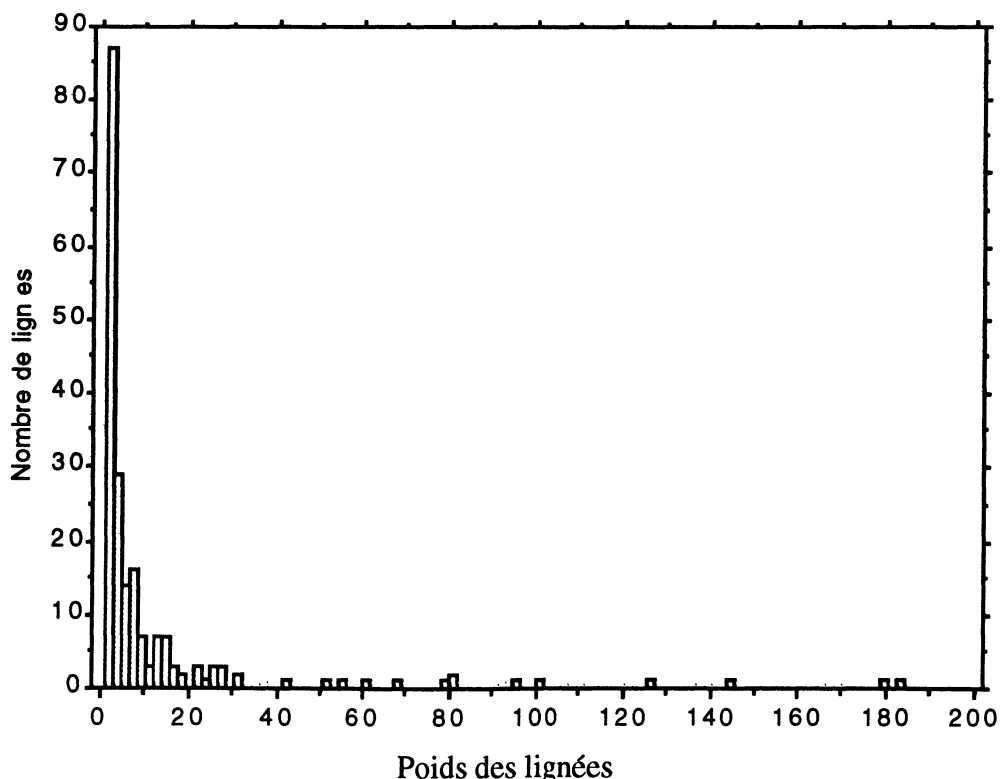


Figure 4 - Représentation de la fréquence des poids bruts

Nous formons donc l'hypothèse que le nombre de sommets de poids x est une fonction de x du type:

$$N(x) = K/x.$$

Une transformation de x en $1/x$ permet de calculer la corrélation de $N(x)$ par rapport à $1/x$ et d'obtenir une estimation de la distribution de p .

$$N(x) = 61/x \quad R^2 = 0,957 \quad R = 0,98$$

Ceci paraît un excellent ajustement mais est, en réalité, obtenu artificiellement par les 3 valeurs élevées de x . L'ajustement sur le nuage de points pour $x < 21$ donne $R = 0,872$. Mais le fait que l'on ait pour des valeurs différentes des poids la même fréquence jette un discrédit sur cette façon de représenter la distribution de p .

Pour cette raison, nous étudions l'hypothèse d'une loi de Pareto-Levy⁹ qui présente de nombreux avantages par rapport aux lois hyperboliques, estimées par passage au logarithme. En effet, la possibilité d'utiliser les propriétés linéaires des valeurs centrales conditionnelles, en général la moyenne ou la médiane, évite la transformation logarithmique et la réduction des écarts qu'elle implique¹⁰.

4.1.1. Rappel sur les distributions de Pareto-Levy

Soient $p_1, p_2, \dots, p_i, \dots, p_n$ les poids bruts observés, rangés par ordre décroissant et $g(p_i)$ la fréquence observée de chacun de ces poids. On pose :

$$G(p_i) = \sum_{j=1}^{j=i} g(p_j)$$

$$m(p_i) = \frac{1}{G(p_i)} \sum_{j=1}^{j=i} g(p_j) p_j$$

Ainsi, $m(p_0)$ est la moyenne des poids de lignée dont le poids est supérieur à p_0 . $m(p_0)$ est la moyenne de tous les poids observés.

G et m sont des estimations de la fonction de répartition $P(p_i)$ et de la moyenne conditionnelle $M(p_i)$.

1) Si $m(x_i)$ est bien représentée par la droite $m(x) = Ax + B$, alors la fonction $P(x)$ suit une des lois suivantes selon les valeurs de A et B .

2) $A > 1, B \neq 0$: Pareto, cas général

$$\frac{K}{(x+c)^\alpha}$$

3) $A = 1, B > 0$: Exponentielle

$$e^{-\frac{x-a}{B}}$$

4) $A > 1; B = 0$: Pareto - Type 1

$$\left(\frac{a}{x}\right)^\alpha, x \geq a > 0$$

⁹ Les premières études sur les renchéissements à Saint Jean Trolimon négligèrent les lois de Pareto-Levy et se contentèrent d'ajustements approximatifs. Les divers articles, cités en bibliographie, et la patience attentive de Marc Barbut ont permis la nouvelle orientation de cette étude. Certains calculs ont même été vérifiés par lui. Qu'il trouve ici l'expression de ma reconnaissance.

¹⁰ Ce qu'on minimise, en fait, ce sont donc non les écarts absolus (entre <<ajustement>> et <<observations>>) mais les *écarts relatifs* pour les observations originelles. Quand les observations ont des ordres de grandeurs d'une grande étendue un bon ajustement sur les logarithmes peut dissimuler un très mauvais ajustement sur les observations non transformées. Marc Barbut *Des bons et des moins bons usages...*

5) $0 < A < 1$; $B > 0$: Distribution Puissance $\left(\frac{x_0 - x}{x_0 - a}\right)^\gamma$, $\gamma = \frac{A}{1 - A} > 0$

6) $A = 1$, $B = 0$: Distribution égalitaire $P(x) = 1$ si $x \leq a$,
 $P(x) = 0$ si $x > a$

4.1.2 Ajustement à une distribution de Pareto-Levy pour la ville de Nanterre Étudions d'abord la ville de Nanterre.

Le tracé de la fonction de p_i , $m(p_i)$, pour la commune de Nanterre montre une linéarisation médiocre. Sur l'intervalle complet des poids, la droite des moindres carrés est estimée par l'équation

$$y = 0,943x + 37,4 \quad \text{avec } R^2 = 0,913 \quad (R = 0,955).$$

Nous avons représenté ci-après la moyenne conditionnelle des poids bruts pour la ville de Nanterre, en visualisant le nombre d'observations par la taille des points représentant ces observations. On vérifie ainsi l'influence des poids faibles (< 40) par rapport aux poids élevés, plus spectaculaires.

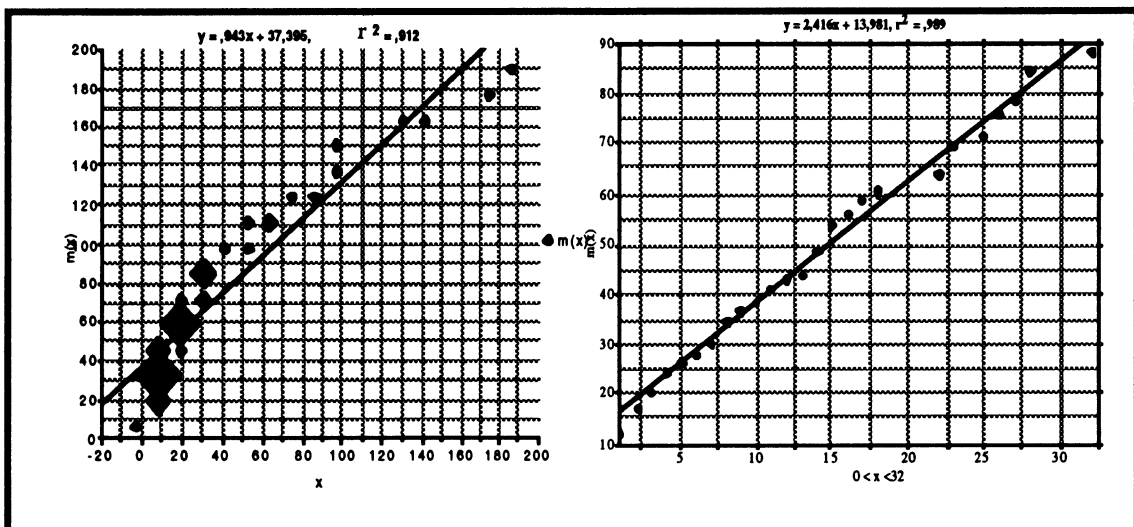


Figure 6. Nanterre - Moyenne conditionnelle des poids

Les coefficients de cette droite ($y = Ax + B$) sont très sensibles aux intervalles de poids choisis. Ainsi le résultat de l'ajustement pour $32 < p < 180$ est très différent mais ne permet pas de conclure positivement à une distribution de Pareto. L'examen plus attentif de l'ensemble des observations montre que 93% des poids observés sont inférieurs ou égaux à 32. L'étude de l'ajustement doit donc être faite pour $1 \leq p_i \leq 32$.

On constate alors un ajustement excellent. Le coefficient de corrélation pour le calcul de la moyenne conditionnelle est: $R = 0,995$ pour une moyenne conditionnelle estimée à :

$$m(p) = 2,4p + 14$$

Ce qui conduit à une estimation de la fonction de répartition :

$$\hat{G}(p) = \frac{K}{(p+10)^{1,714}}$$

Le coefficient K peut être estimé de plusieurs façons :

- en minimisant la somme des valeurs absolues de $\hat{G}(p) - G(p)$ sur l'intervalle $2 \leq p_i \leq 180$.
- en minimisant les valeurs extrêmes sur lesquelles se construit le test de Kolmogorov¹¹

Dans le cas de Nanterre la première méthode donne $K=9088$, la deuxième $K = 9300$. Pour cette valeur de K , les valeurs extrêmes de $\delta = \hat{G}(p) - G(p)$ sont de $-7,8$ ($x = 2$) et $+7,8$ ($x = 1$), qui ramenées au nombre d'observations ($N = 201$), fournissent un écart maximum de 3,9%. Le test de Kolmogorov conclut largement à l'acceptation de l'hypothèse d'une distribution parétienne sur l'intervalle $2 \leq p_{ii} \leq 180$.

4.1.3. Ajustement à une distribution de Pareto-Levy pour la commune de Saint Jean Trolimon
Le calcul de la fonction de répartition donne des résultats similaires à ceux de Nanterre. La droite des moindres carrés permettant l'estimation de la moyenne conditionnelle est obtenue avec $R^2 = 0,934$ (valeur faible qui conduit à $R = 0,97$) pour tout l'intervalle des poids.

Le tracé de $m(p)$ montre aussi des discontinuités dans l'intervalle d'étude qui est ici de $p = 1$ à $p = 300$. La première apparaissait autour de $p = 32$ comme à Nanterre. Une seconde peut être relevée autour de $p = 150$. Elle était déjà amorcée dans le cas de Nanterre mais ne concernait pas assez de sommets pour être prise en compte. La pente de la droite correspondant à l'intervalle complet est proche de 1 (1,024) comme elle l'était pour Nanterre(0,93). Dans un cas comme dans l'autre on aura une mauvaise estimation sur l'ensemble de l'intervalle en ajustant une loi de Pareto exponentielle $Ke^{-\frac{x}{B}}$.

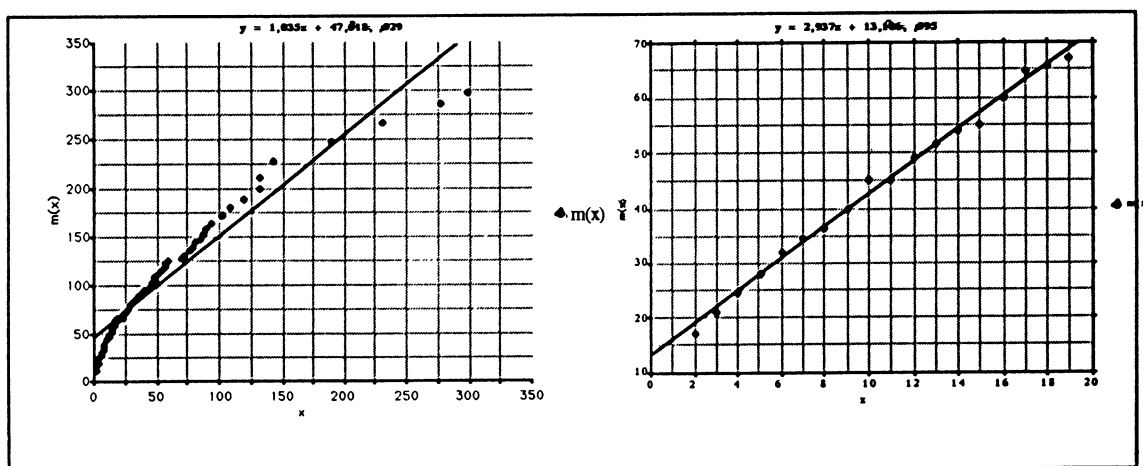


Figure 6. Saint Jean Trolimon - Moyenne conditionnelle des poids

¹¹ Plusieurs vérifications peuvent être appliquées pour s'assurer de l'admissibilité d'une représentation Parétienne : l'écart relatif moyen $\hat{G}(p) - G(p)$ ramené au nombre d'observations, le test de Kolmogorov et Smirnov porte, lui, sur l'écart maximum. On peut aussi calculer l'indice d'inégalité de Gini et le comparer à sa valeur calculée à partir des données. Marc Barbut, dans l'article précédemment cité, conseille aussi le contrôle par la médiane.

Intervalle des poids pour le calcul de la moyenne conditionnelle	Fonction	Pourcentage des observations dans l'intervalle	Ecart maximum entre valeurs observées et calculées
2 à 20	$G(p) = \frac{7370}{(p + 6,74)^{1,5}}$	87%	< 4,9% pour $p = 2$ < 2,6% pour $p = 6$
20 à 150	$G(p) = \frac{K}{(p + 167)^{4,8}}$	12%	Non calculé
150 à 300	non calculé	1%	Non calculé
2 à 150	$G(p) = \frac{9 \times 10^8}{(p + 88,7)^{3,85}}$	99%	< 6% pour $p > 7$

On vérifie que l'intervalle 2-20, concernant 87% des observations donne lieu à un ajustement excellent ($R^2 = 0,995$, $R = 0,997$) avec la distribution de Pareto

$$G(p) = \frac{7370}{(p + 6,74)^{1,5}}$$

sauf pour les valeurs extrêmes $p = 1$, $p = 2$. L'écart maximal $\Delta = \hat{G}(p) - G(p) / 426$ est inférieur à 2,7% et à 4,9% pour $x = 2$. On vérifie que la prise en compte d'un intervalle plus important pour le calcul de la moyenne conditionnelle $1 < p < 151$, concernant 99% des observations donne un bon ajustement avec la distribution de Pareto. $G(p) = \frac{900.000.000}{(p + 88,7)^{3,85}}$

mais uniquement pour les poids supérieurs ou égaux à 6 ; choisir cet ajustement c'est ignorer les poids les plus fréquents¹². Même sur l'intervalle restreint $p > 7$, les valeurs extrêmes de Δ sont proches de 6%. Nous considérerons donc que la fonction de répartition des poids bruts pour Saint Jean Trolimon est bien estimée par :

$$G(p) = \frac{7370}{(p + 6,74)^{1,5}}$$

4.1.4 Comparaison des distribution de poids entre Nanterre et Saint Jean Trolimon

On rappelle dans le tableau ci-joint les paramètres essentiels de cette comparaison.

Valeurs comparées	NANTERRE	SAINT JEAN TROLIMON
Taille du fichier	4426	5078
Nombre de couples repérés	4408	5356
Nombre d'ancêtres pertinents	200	426
Nombre de renchaînements	1214	2359
Intervalle des poids bruts	1 - 183	1 - 299
Moyenne des poids bruts	12,15	12,7
Somme des poids bruts	2.430	5.408
Fonction de répartition	$G(p) = \frac{9300}{(p + 10)^{1,7}}$	$G(p) = \frac{7370}{(p + 6,74)^{1,5}}$
Maximum $\hat{G}(p) - G(p)$ %	3,9%	2,7%
Moyenne conditionnelle	$y = 2,416x + 13,981$ $R = 0,995$	$y = 2,937x + 13,106$ $R = 0,997$

¹² Les désaccords entre les valeurs observées et calculées sont souvent élevés pour les valeurs faibles, ce qui est habituel dans le cas des fonctions de Pareto. Néanmoins, il serait regrettable de choisir un ajustement qui ne prendrait pas bien en compte les valeurs les plus représentées.

Comparer une quelconque propension au renchaînement entre ces deux populations c'est, soit estimer que les différences de taille d'échantillon ou de mode de construction de cet échantillon sont des facteurs non pertinents, soit se donner les moyens de construire une mesure qui prend en compte ces différences. Nous constatons que :

Nanterre produit 2 fois moins de renchaînements que Saint Jean Trolimon et 2 fois moins d'observations. Le poids le plus élevé à Nanterre est de 183, il est de 299 à Saint Jean Trolimon. Par contre, les moyennes sont très proches. Donc en première analyse, ce que ces populations donnent à voir de leur aptitude à renchaîner à partir de l'échantillon choisi n'est pas très discriminant. Mais l'étude des moyennes conditionnelles et des distributions de Pareto permet d'affiner ces conclusions. Sur les intervalles où elles ont été estimées les moyennes conditionnelles sont bien représentées par les droites :

$$\begin{array}{ll} \text{Nanterre} & m = 2,416p + 13,98 \\ \text{Saint Jean Trolimon} & m = 2,94p + 13,401 \end{array}$$

Les différences, mises en évidence par les distributions parétiennes, sont faibles. On note toutefois que la moyenne conditionnelle pour Saint Jean Trolimon est légèrement supérieure à celle de Nanterre pour tous poids supérieurs à 1,7. En dessous de ce poids limite, la moyenne de Nanterre est supérieure. Mais il s'agit de différences faibles. En tout cas, on peut en conclure que les distributions de poids bruts sont semblables, bien que Saint Jean Trolimon produise plus de renchaînements d'alliance et des ancêtres à poids plus élevé. Par contre Nanterre produit proportionnellement plus d'ancêtres à poids très faibles.

4.1.5. Comparaison des distributions de poids par la représentation de LORENZ-GINI

Pour rendre plus manifeste encore la comparaison entre ces deux communes, nous avons utilisé la représentation de LORENZ-GINI¹³. Dans cette représentation, on indique en abscisse le pourcentage cumulé des observations, en ordonnée le pourcentage cumulé du produit [poids * observations].

$$x_j = \frac{\sum_{i=n}^{i=j} g(x_i)}{\sum_{i=n}^{i=l} g(x_i)} \quad y_j = \frac{\sum_{i=n}^{i=j} x_i g(x_i)}{\sum_{i=n}^{i=l} x_i g(x_i)}$$

Les courbes de concentration de Gini confirment l'identité relative des deux communes, mesurée en termes de répartition des poids bruts.

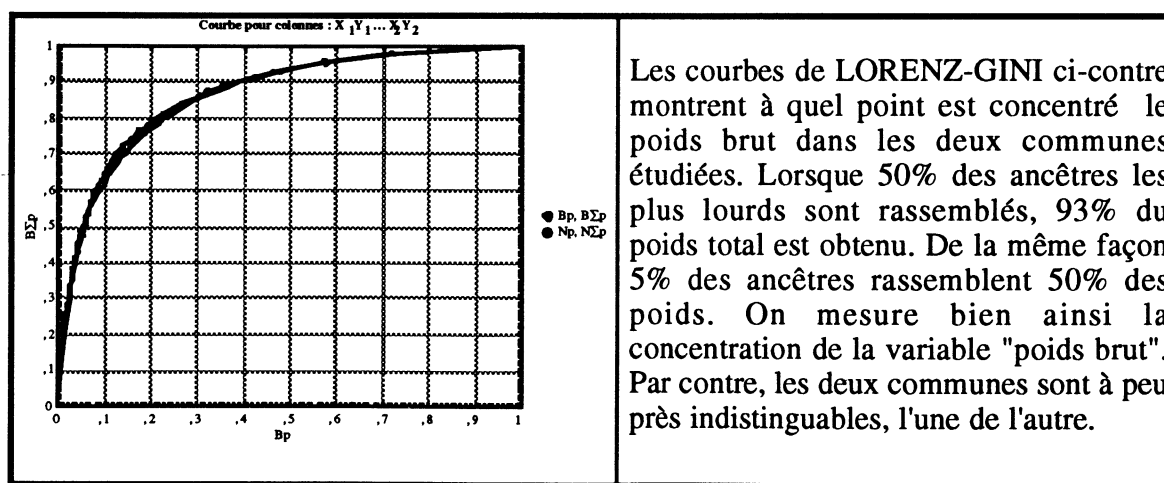


Figure 7. Concentration des poids bruts - Comparaison entre les deux communes

¹³ Bien que l'on puisse déduire les courbes de Gini de l'estimation faite d'une distribution de Pareto, les courbes de Gini ont plusieurs avantages. Elles manifestent clairement le phénomène de concentration des poids et ceci, d'une façon globale. Elles facilitent la comparaison des deux communes, même lorsque l'on connaît les distributions parétiennes correspondantes. La prise en compte simultanée des 3 paramètres K , c , α n'est pas aisée, en particulier quand ils diffèrent beaucoup d'une commune à l'autre.

4.2. Les poids réduits

4.2.1. Construction des poids réduits

Les paramètres des deux populations diffèrent par le nombre de renchaînements d'alliance relevés et par l'intervalle des poids observés (183 contre 299). Mais la répartition des tailles de lignées (nombre de descendants d'un individu) fait apparaître une cause de différence massive. Il est vraisemblable que le nombre de renchaînements d'alliance observés pour un ancêtre donné, croît avec la taille de la lignée de cet ancêtre. Aussi sommes nous conduits à étudier la répartition des tailles de lignées et une nouvelle variable, le poids réduit, construite comme suit. Pour tout ancêtre de poids p non nul, engendrant une lignée de taille l , on appelle "poids réduit" la variable $100 \times \frac{p}{l}$. Le poids réduit permet d'étudier les renchaînements d'alliance sans être sous la dépendance du manque d'informations sur les ancêtres éloignés.

4.2.2. Étude de la répartition des tailles de lignées

Nous avons essayé d'ajuster la répartition des tailles de lignées par une loi de Pareto. La figure ci-dessous présente les moyennes conditionnelles pour les deux communes.

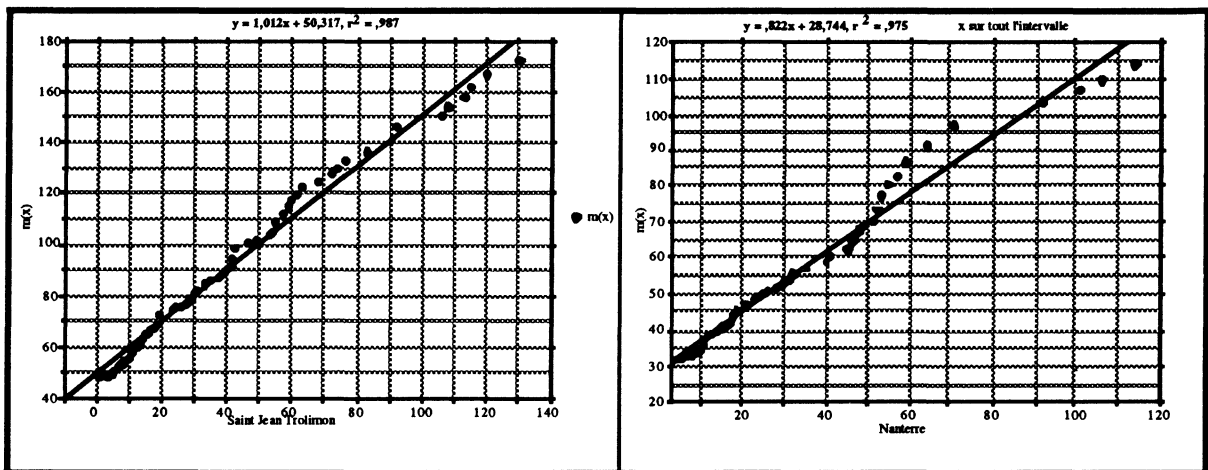


Figure 8. Moyenne conditionnelle pour les tailles de lignées des deux communes

Pour la ville de Nanterre, en choisissant l'intervalle qui maximise r^2 , la moyenne conditionnelle de la taille des lignées est ajustée par la droite

$$m(x) = 0,81 x + 28,3 \quad R = 0,995$$

ce qui, compte tenu du coefficient de x inférieur à 1, donne un ajustement par une fonction puissance. $\hat{G}(x) = K \left(\frac{147 - x}{147 - 3} \right)^{4,21}$; $3 \leq x \leq 147$. La valeur $x_0 = 147$ est au delà de notre intervalle, donc pertinente. La valeur de K peut être ajustée à 74 avec des écarts de $\pm 4,27\%$.

Pour la commune de Saint Jean Trolimon, l'ajustement de la moyenne conditionnelle se fait avec un coefficient de corrélation de 0,995 et conduit à une fonction de répartition

exponentielle $\hat{G}(x) = 94 e^{\frac{-x}{50}}$. Les écarts maxima sont de -6,7 et +6,3 sur tout l'intervalle des tailles et très rapidement plus faibles si l'on élimine les tailles faibles. La représentation ci-après du nombre conditionnel d'observations pour les deux communes montre une différence sensible, à la fois dans l'étendue de l'intervalle des tailles et dans le nombre d'observations. La représentation par les courbes de Lorenz-Gini manifeste une concentration supérieure pour la commune de Saint Jean Trolimon. En effet 20% d'observations correspondent à 44% du total des lignées à Nanterre, 56% à Saint Jean Trolimon.

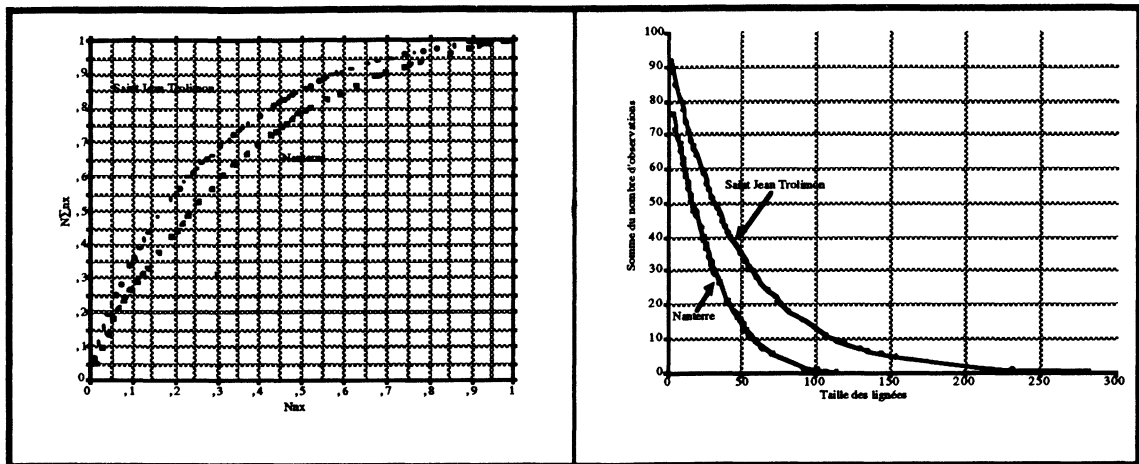


Figure 9. Représentation des tailles de lignées des deux communes

L'étude des poids devrait prendre en compte ces différences en construisant une variable qui soit le plus indépendante possible de l'échantillonnage.

4.2.3. Étude des poids réduits

La comparaison des répartitions des poids réduits est exprimée dans le tableau ci-après.

Valeurs comparées	NANTERRE	SAINT JEAN TROLIMON
Taille du fichier	4426	5078
Nombre de couples repérés	4408	5356
Nombre d'ancêtres pertinents	200	426
Nombre de renchaînements	1214	2359
Intervalle des poids réduits	2-257	1-393
Moyenne des poids réduits	40,7	28,7
Somme des poids réduits	8186	12157
Fonction de répartition	$\hat{G}(x) = 213 e^{\frac{-x}{41}}$	$\hat{G}(x) = \frac{106.10^5}{(x + 42,7)^{2,7}}$
Maximum $\hat{G}(p) - G(p)$ %	-7,9% - +8,5%; $x > 1$	-2% - +3% ; $x > 1$
Moyenne conditionnelle	$y = 1,02x + 39,3$ $R = 0,992$	$y = 1,6x + 25,7$ $R = 0,996$

L'intervalle des poids réduits reste dans le même rapport que celui des poids bruts. Mais la différence sur la moyenne devient significative. En moyenne, il y a 41 renchaînements d'alliance dans une lignée, pour 100 descendants à Nanterre. Il y en a 28 à Saint Jean Trolimon. Mais la moyenne conditionnelle croît plus vite à Saint Jean Trolimon : elles deviendront égales pour un poids réduit de 23%, correspondant à une moyenne de 63%.

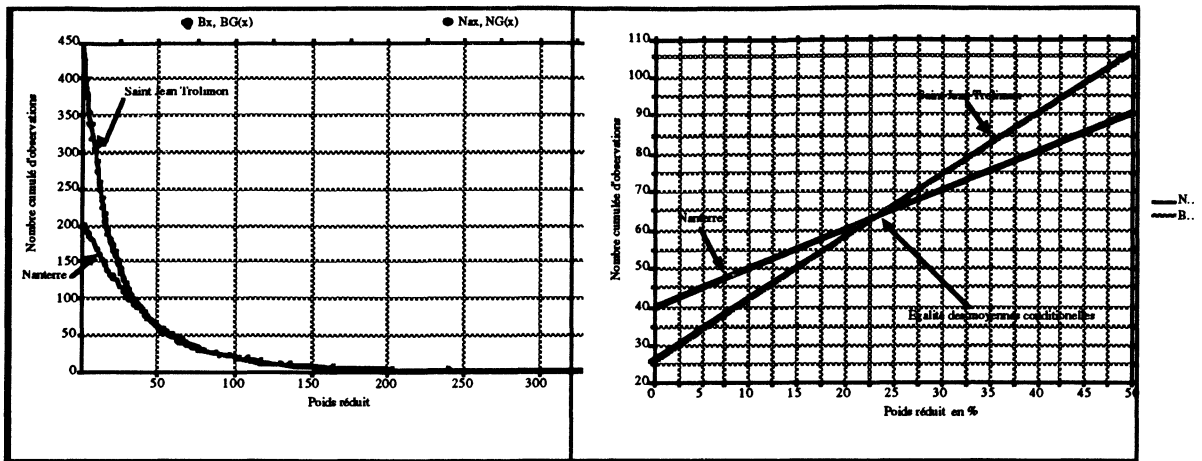


Figure 10. Représentation des poids réduits des deux communes.

Saint Jean Trolimon produit des poids réduits élevés, de 400 à 260, en nombre faible. A partir de 260, jusqu'à 30, Nanterre et Saint Jean Trolimon sont équivalents malgré le déficit de Nanterre en nombre de renchaînements d'alliance. De 30 à 1, Saint Jean Trolimon cumule un beaucoup plus grand nombre d'observations que Nanterre, mais, en moyenne, Nanterre est plus productif dans cet intervalle. Les courbes de concentration de Lorenz-Gini confirment cette interprétation. Saint Jean Trolimon est plus concentré que Nanterre, il produit plus de lignées extrêmes, de poids réduits élevés ou très faibles. 25% des lignées de poids réduits élevés correspondent à 50% du total des poids à Nanterre, 60% à Saint Jean Trolimon.

En conclusion provisoire nous pouvons dire que la détermination des distributions parentes et l'analyse des moyennes conditionnelles permet de distinguer clairement les comportements des deux communes en matière de renchaînements.

5. LA DISTRIBUTION DES COUPLES RENCHAÎNANTS

Le renchaînement d'alliance met en relation deux couples. Chacun de ceux-ci est en relation avec d'autres couples dans d'autres renchaînements d'alliance. Chaque couple est donc associé à un ou plusieurs autres couples. Nous appellerons "surface d'un couple" le nombre de couples qui lui est associé. Les échantillons dont nous disposons peuvent être étudiés du point de vue de la répartition des surfaces de couples renchaînant au moins une fois. Bien entendu la surface d'un couple repéré dans un échantillon dépend de la profondeur connue des généalogies. On peut donc s'attendre à une surface moyenne plus faible à Nanterre.

5.1. Construction des distributions parentes

L'analyse rapide des distributions de fréquences montre que, comme dans l'étude des poids, le logarithme du nombre cumulé d'observations est une fonction linéaire de la surface. Nous essaierons donc un ajustement à une distribution de Pareto. Le tableau ci-joint résume les résultats obtenus.

Taille du fichier	4426	5078
Nombre de couples repérés	4408	5356
Nombre de couples pertinents	357	659
Nombre de renchaînements	1214	2359
Intervalle des surfaces	1 - 51	1 - 76
Moyenne des surfaces	6,81	8,21
Moyenne conditionnelle	$y = 0,954x + 8$ $R = 0,995$	$y = 0,953x + 12,7$ $R = 0,991$
Moyenne de $m(x) - x$	7,2	11,6
Ajustement à l'exponentielle	$\hat{G}(x) = 330 e^{\frac{-x}{7,2}}$	$\hat{G}(x) = 486 e^{\frac{-x}{7,2}} e^{\frac{-x}{11,2}}$
Maximum $\hat{G}(p) - G(p)$ %	-6% - +6%; $x > 1$	-7% - +7% ; $x > 2$

On notera d'abord la différence entre la moyenne calculée sur l'ensemble des observations (6,8 / 8,2) et celle qui est obtenue à partir de la droite représentant la moyenne conditionnelle. Comme cela est souvent le cas, les valeurs extrêmes (Surface de 1 ou 2) sont mal représentées par les distributions de Pareto. En deuxième lieu, nous constatons que les deux distributions parentes sont exponentielles, avec des coefficients en x de la moyenne proportionnelle très semblables, très proches de 1, mais en différant légèrement. Ceci explique que les écarts $\hat{G}(x) - G(x)$ sont, pour les deux communes, de l'ordre de 6% pour des surfaces de couples supérieures à 2. Peut-on trouver dans les combinatoires qui président aux renchaînements d'alliance une explication au fait de reconnaître, pour deux populations si différentes, des distributions exponentielles, nous ne pouvons le dire.

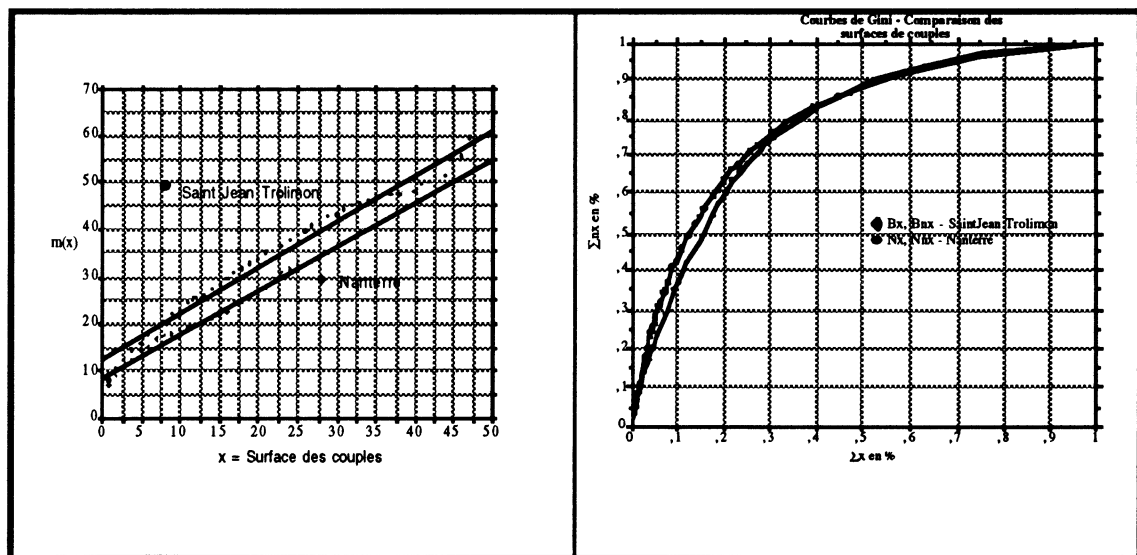


Figure 11. Représentation des surfaces des deux communes

5.2. Comparaison par la courbe de LORENZ-GINI

Les courbes de concentration de Lorenz-Gini montrent, une fois de plus, que Saint Jean Trolimon est plus concentré que Nanterre : 20% des observations correspondent à 48% de la surface cumulée pour Nanterre, 56% pour Saint Jean Trolimon. Ces différences ne sont pas considérables et ont de quoi étonner quand on connaît les différences manifestées par les deux échantillons en terme de profondeur généalogique. D'ailleurs les coefficients de Gini, calculés sur les distributions empiriques, sont de 0,298 à Nanterre et de 0,311 à Saint Jean Trolimon.

5.3. Influence de la connaissance des généalogies

Nous avons essayé de mettre en parallèle le rapport des moyennes $7,2 / 11,6 = 60\%$ et divers autres coefficients construits à partir des statistiques des populations.

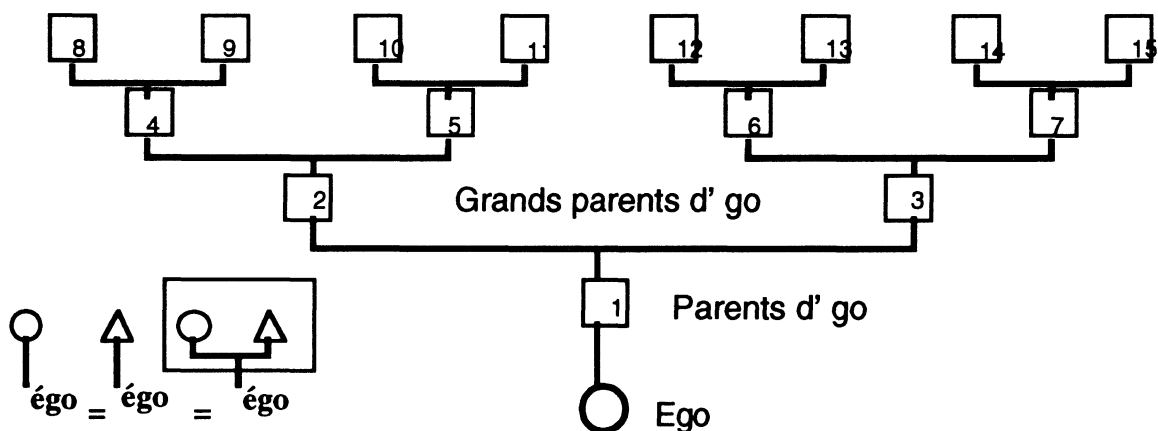
- 1) Nombre de couples concernés.
- 2) Surface maximale.
- 3) Taille de la population.

Aucun de ces coefficients ne semble plus justifié qu'un autre pour redresser les effets pervers dus à la sous-représentation de certains individus dans notre échantillon.

Certes, la probabilité qu'un renchaînement entre deux couples a et b soit repérée croît avec l'information généalogique sur a et sur b. La surface observée d'un couple est d'autant plus proche de la surface réelle que le nombre de parents connus est plus proche du nombre maximum compatible avec le niveau de profondeur maximum prévu pour définir le concept de renchaînement. (4 degrés = 15 couples = 30 individus). Il serait donc intéressant de pondérer les surfaces et en particulier les surfaces moyennes par un coefficient représentant l'écart à 15 du nombre de couples ancêtres connus ou l'écart à 30 du nombre d'individus-ancêtres. C'est cette mesure de la connaissance généalogique que nous allons aborder maintenant.

5.4. Mesure de la connaissance généalogique

1) Remarquons d'abord que cette mesure, apparemment simple ne fournit pas un coefficient indiscutable. En effet la connaissance d'un ascendant sur deux dans la dernière génération connue est suffisante pour repérer les renchaînements vrais (par opposition aux renchaînements de type demi ; par l'intermédiaire d'un demi-germain). On repérera donc la présence d'un couple ancêtre par la présence d'un de ses membres.



Calcul du Coefficient de Connaissance Généalogique

Figure 12. La connaissance généalogique

2) Deux généalogies disposant du même coefficient γ , l'une par connaissance des 4 ancêtres paternels jusqu'à la quatrième génération et d'aucun ancêtre maternel, (du père à l'arrière-arrière-grand-père), l'autre par le couple des parents, des arrière-grands-parents et d'un ancêtre d'arrière-grands-parents, ne sont pas équivalentes du point de vue du renchaînement. Car les renchaînements proches étant plus fréquents que les renchaînements lointains, une perte d'information de bas niveau généalogique est plus importante que son équivalent dans les niveaux élevés.

Il faudrait donc raffiner le coefficient γ en faisant dépendre de l'absence de l'ancêtre le poids calculé mais en tenant compte pour cette absence:

- de la présence de l'autre conjoint,
- de la profondeur généalogique où elle se situe¹⁴.

Dans un premier temps nous en resterons à un coefficient simple en ne tenant compte que de la première remarque et donnerons à γ la définition suivante¹⁵ :

$$\gamma = \frac{\sum_{i=1}^n \frac{\min(a(x_i), 30)}{30}}{n}$$

avec n = Nombre d'individus connus et $a(x_i)$ = nombre d'ancêtres de x_i connus dans l'échantillon.

La moyenne d'ancêtres connus est alors pour Nanterre de 3,34 , pour Saint Jean Trolimon de 9,17. On vérifie, une fois de plus que, malgré son handicap pour le repérage des renchaînements, Nanterre (rapport $3,3/9,2 = 36\%$) a une moyenne de surface de couples de 60% par rapport à Saint Jean Trolimon. Donc Nanterre est beaucoup plus productrice de renchaînements que Saint Jean Trolimon. C'est une conclusion qualitative et provisoire car il paraît osé de définir une moyenne de la surface des couples rapportée au nombre moyen d'ancêtres connus soit $7,2/3,3 = 2,2$ pour Nanterre et $11,6/9,2 = 1,3$ pour Saint Jean Trolimon. Une telle mesure nous indiquerait que Nanterre est deux plus fois plus renchaîneuse que Saint Jean Trolimon. Il importe de contrôler ce rapport de façon plus précise.

5.5. Étude des conditions du repérage d'un renchaînement d'alliance

Soit la figure formelle de renchaînements d'alliance (Fig 1). Pour qu'un tel Renchaînements d'alliance soit repéré il faut que soient connus:

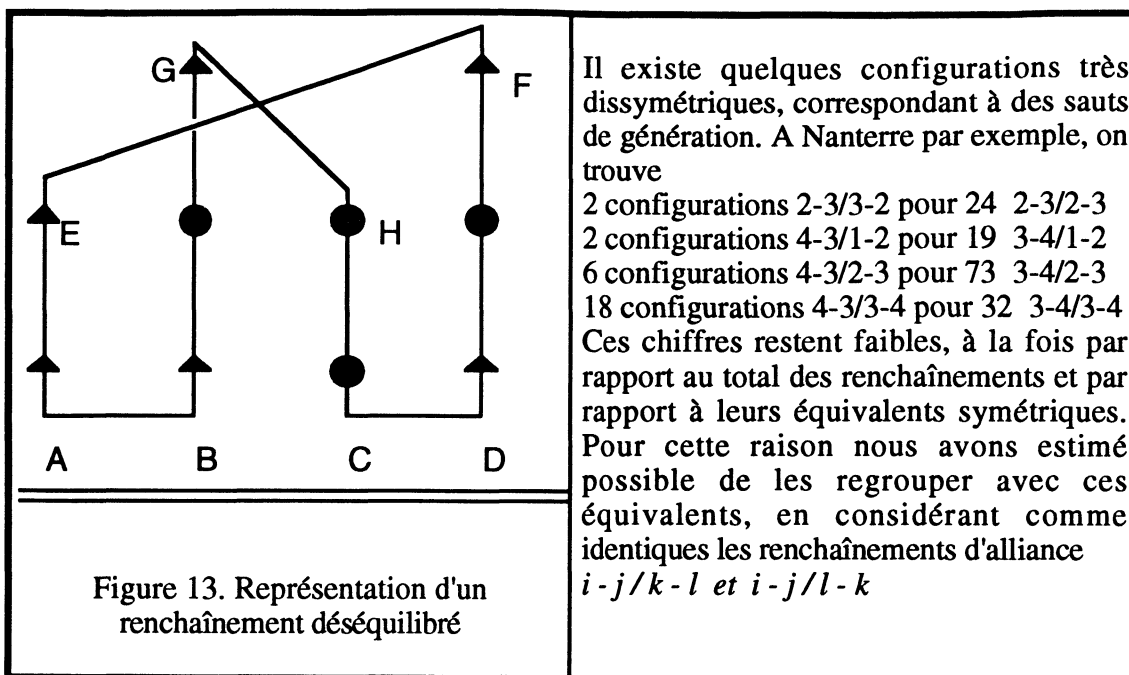
- le mariage AB
- le mariage CD
- la consanguinité AMC
- la consanguinité BND

Nous partons de l'hypothèse (vérifiée à Nanterre en tout cas, moins certainement à Saint Jean Trolimon) que les mariages sont tous connus. Si ce n'était pas le cas, il faudrait faire intervenir un facteur correctif. Par contre, les ancêtres M et N ne sont repérés que si les généalogies sont connues jusqu'au niveau de M et de N. Un lien de consanguinité AMC, compte tenu de notre hypothèse de recherche, peut prendre toutes les valeurs de 1-1 (A et C germains) à 4-4 (A et C cousins au 3^è degré : M arrière-arrière-grand-père). Soient les configurations : 1-1, 1-2, 1-3, 1-4, 2-2, 2-3, 2-4, 3-3, 3-4, 4-4. La figure de renchaînement est caractérisée par un 4-uplet: (i, j, k, l) avec

$$\begin{aligned} 1 &\leq i, j, k, l \leq 4 \\ j &\geq i ; l \geq k \\ i &\geq k ; j \geq l \end{aligned}$$

¹⁴ Le poids réduit permet, à la fois, de tenir compte de la profondeur généalogique (nombre de générations entre l'ancêtre et son descendant le plus lointain) et du nombre moyen de descendants à chaque génération.

¹⁵ Ce coefficient généralise l'idée intuitive suivante : un individu nous permet de retrouver les renchaînements qui le concernent s'il a 30 ancêtres connus. Au delà de 30, les ancêtres connus ne nous informent pas. Si l'individu n'a que 15 ancêtres connus, nous risquons de ne retrouver que la moitié des renchaînements. Une société dans laquelle on serait informé sur les 30 ancêtres de chaque individu n'existe pas par suite des bornes de l'échantillon. Mais on peut construire un sous-ensemble dans lequel $\gamma = 1$. Par ailleurs, l'avantage de ce coefficient réside dans sa capacité à exprimer globalement la connaissance généalogique que nous avons de la société.



S'autoriser une telle confusion, par exemple, 2-1 et 1-2 ou 1-3 et 3-1, c'est être cohérent avec le type d'analyse qui est proposé en 5.6 et avec le concept de couple efficace pour lequel les deux conjoints ne sont pas différenciés.

Nous avons repéré la représentation des 50 figures de renchaînements d'alliance selon les valeurs possibles de ces 4-uplets. Cette représentation est donnée ci-contre. On peut y lire que, à Nanterre, il existe 20 renchaînements 1-1/1-1, c'est-à-dire entre germains des deux côtés et 34 renchaînements d'alliance entre cousins au 3ème degré.

Type liens	1-1	1-2	1-3	1-4	2-2	2-3	2-4	3-3	3-4	4-4	Total	%
1-1	20	0	0	0	22	2	0	26	10	14	114	4,7%
1-2		21	8	0	6	54	9	20	21	11	171	7,0%
1-3			19	14	0	15	21	0	6	0	102	4,2%
1-4				5	0	1	9	0	0	0	34	1,4%
2-2					21	35	2	56	43	28	234	9,6%
2-3						46	26	53	83	36	397	16%
2-4							18	5	15	1	124	5,1%
3-3								57	108	93	475	20%
3-4									50	70	456	19%
4-4										34	321	13%
											2428	

Nanterre - Répartition des renchaînements bruts selon les types de liens

Type liens	1-1	1-2	1-3	1-4	2-2	2-3	2-4	3-3	3-4	4-4	Total	%
1-1	52	22	0	0	24	32	3	44	37	52	318	5,9%
1-2		30	13	3	24	80	25	41	103	39	410	7,6%
1-3			28	33	4	42	79	6	67	5	305	5,6%
1-4				31	0	3	34	0	5	0	140	2,6%
2-2					24	46	9	51	63	63	332	6,1%
2-3						75	64	92	178	100	787	15%
2-4							61	13	121	16	486	9,0%
3-3								51	148	103	600	11%
3-4									193	158	1266	23%
4-4										114	764	14%
											5408	

Saint Jean Trolimon - Répartition des renchaînements bruts selon les types de liens

Nous avons marqué en gras les figures de renchaînements qui se réalisent sans décalage de génération entre conjoints : les plus fréquents pour les liens 1 et 2. Au delà, le nombre de générations (i. e. la taille du lien) est assez grand pour faciliter les réajustements. Ainsi en est-il de la configuration 3-4/4-4, par exemple.

Nous avons calculé le pourcentage par rapport au nombre total de renchaînements, des renchaînements dont un des liens reste constant. Soit, tous les 1-1/*i-j*, les 1-2/*i-j*...etc. On vérifie que les liens symétriques sont privilégiés et que le nombre brut de renchaînements d'alliance croît avec la longueur des liens. Il y a plus de liens 4-4 que de liens 1-1. Il y a plus de liens 4-4 que de liens 1-4. La sommation de toutes les configurations ayant au moins une relation 1, une relation 2, .. etc. est présentée ci-dessous :

Liens	Nombre brut de renchaînements				Pourcentage par rapport au total			
	1	2	3	4	1	2	3	4
Nanterre	421	755	931	321	17%	31%	38%	13%
Saint Jean	1163	1605	1866	764	22%	30%	35%	14%

On constate que, vue sous l'angle des pourcentages, la structure des renchaînements est indépendante des populations concernées et de leur aptitude respective à renchaîner. On peut supposer que l'on subit, ici, les effets conjugués de la combinatoire des généalogies (nombre de conjoints disponibles à chaque niveau) et de la connaissance plus ou moins grande des ancêtres dans les deux populations. Mais cette conclusion est toute provisoire.

Le nombre de renchaînements d'alliance observés est, en réalité, un échantillon du nombre réel dans la population. Cet échantillon est composé des paires de couples pour lesquels les ancêtres de niveau *i-j* (couple 1) et *k-l* (couple 2) sont connus. L'estimation du nombre réel de renchaînements d'alliance peut être faite à partir des propriétés des échantillons ainsi définis et met en jeu la notion de *COUPLE EFFICACE* et d'*EQUIVALENT-COUPLE* que nous voyons maintenant.

5.6. La notion de couple efficace

Un couple est efficace pour un renchaînement d'alliance de type *i-j / k-l* si l'on connaît les ancêtres de ce couple permettant de repérer le renchaînement. Pour un renchaînement d'alliance

1-1/ $k-l$, il faut connaître le couple parental. Dans un premier travail¹⁶ nous calculions le nombre de couples efficaces par une mesure globale du nombre d'ancêtres connus qui nous amenait à conclure en la présence d'un ancêtre au niveau 1, 2 ou 3. Cette méthode ne tenait pas compte de la répartition des ancêtres connus (côté paternel ou maternel, par exemple). Mais surtout, elle ignorait les différences entre les individus ayant de 1 à 32 ancêtres à la 4ème génération, de 1 à 16 à la 3ème, etc.. En effet, un couple peut être complètement efficace pour un niveau donné si il possède tous ses ancêtres ou seulement à moitié efficace, s'il n'en possède que la moitié. Cette notion intuitive peut être formalisée comme suit.

Un individu est i -efficace avec le coefficient d'efficacité e_i si, le nombre de couples parents à la génération i étant n_i on pose : $e_i = \frac{n_i}{2^{i-1}}$. Une 1-efficacité peut prendre les valeurs 0,1; une 3-efficacité peut, elle, prendre les valeurs 0,1/4, 1/2, 3/4, 1. La notion de i -efficacité permet de définir, pour tout individu, son efficacité $e = \{e_1, e_2, e_3, e_4\}$. Cette notion peut être étendue à un couple. Un couple est $i-j$ efficace avec le coefficient d'efficacité e_{i-j} s'il est composé de deux individus i -efficace et j -efficace. On posera donc : $e_{i-j} = e_i \cdot e_j = \frac{n_i}{2^{i-1}} \cdot \frac{n_j}{2^{j-1}}$

Étant donné un nombre de renchaînements d'alliance donné, de type $i-j/k-l$, obtenu dans une population dont tous les ancêtres sont connus (id est : tous les couples ont une $i-j$ -efficacité de 1), le pourcentage de renchaînements d'alliance de ce type, dans cette population, (nombre de renchaînements d'alliance réduit à la population) serait obtenu par division par le nombre total de couples dans la population. En réalité, de nombreux couples ne participant pas au repérage des renchaînements d'alliance, par manque d'information sur leur généalogie ascendante, il importe de compter seulement les couples efficaces en tenant compte de leur efficacité. D'où la notion d'équivalent-couple pour un lien $i-j$.

$C_1, C_2, \dots, C_t, \dots, C_n$ étant l'ensemble des couples formés dans une population produisant n couples, on appelle équivalent-couple de type $i-j$ la quantité $E_{i-j} = \sum_{i=1}^n e_{i-j} (C_i)$. Le nombre réduit de renchaînements d'alliance peut alors s'en déduire comme suit : $r_{i-j/k-l}$ étant le nombre observé de renchaînements d'alliance pour la configuration $i-j/k-l$, $\bar{r}_{i-j/k-l} = \frac{r_{i-j/k-l}}{E_{i-j} \times E_{k-l}}$

On présente ci-dessous les valeurs des équivalents-couples pour les commune de Nanterre et de Saint Jean Trolimon et les renchaînements réduits qui peuvent en être déduits.

	Nanterre				Saint Jean Trolimon			
Liens	1	2	3	4	1	2	3	4
1	766	420	192	49	1434	936	590	357
2		327	163	42		710	472	291
3			121	36			358	235
4				20				170

Ainsi Nanterre produit 766 équivalents-couples de type 1-1, c'est à dire dont les parents des deux membres sont connus. Saint Jean Trolimon peut en présenter 1434, soit deux fois plus. Ce rapport est encore plus en faveur de Saint Jean Trolimon pour le lien 4-4, soit 8,5.

¹⁶ Il s'agissait d'une généralisation de la notion de poids réduit. Nous rapportons le nombre de renchaînements au nombre de couples capables de réaliser ces renchaînements. Le problème est toujours de déterminer ces couples et de les compter. La méthode présentée ici est plus rigoureuse, mais échoue à tenir compte des renchaînements dissymétriques.

A chaque population correspond donc une séquence de 10 nombres qui sont les équivalents-couples pour les dix configurations 1-1, 1-2, ... , 3-3, 3-4, 4-4. Deux ensembles disjoints de N_1 et N_2 couples dans lesquels tous les renchaînements possibles sont réalisés produiront $(N_1 - N_2) / 2$ renchaînements qui seront repérés si les ancêtres nécessaires sont connus. Un couple ne renchaînant pas avec lui-même, c'est $N_1 - (N_1 - 1) / 2$ qui sont produits sur des ensembles non disjoints. En rapportant, pour chaque type de renchaînements, le nombre observé au nombre maximum possible, calculé sur les équivalents-couples, on se donne une mesure largement indépendante de la connaissance que l'on a de la population.

Les tableaux suivants présentent les nombres réduits de renchaînements en tenant compte de ces équivalents-couples. Ceux-ci sont rappelés dans les lignes et colonnes marginales. Les nombres réduits sont ramenés à 100-100, c'est à dire, qu'ils sont le nombre de renchaînements d'alliance que produiraient deux ensembles de 100 couples d'efficacité 1 pour les relations considérées. Les valeurs de N étant grandes, on a ramené $N-1$ à N .

Nanterre - Nombres réduits											
Type liens	1-1	1-2	1-3	1-4	2-2	2-3	2-4	3-3	3-4	4-4	Couples efficaces
1-1	0,7	0	0	0	1,8	0,4	0	5,6	7,2	18,2	766
1-2		2,4	2	0	0,8	15,8	10,2	7,8	27,8	26,2	420
1-3			10,4	30	0	9,6	52	0	17,4	0	192
1-4				41,6	0	2,6	87	0	0	0	49
2-2					4	13,2	3	28,4	73	86	327
2-3						35	76	54	283	220	163
2-4							204	20	198	24	42
3-3								78	495	768	121
3-4									771	1944	36
4-4										1700	20
	766	420	192	49	327	163	42	121	36	20	

Saint Jean Trolimon - Nombres réduits											
Type liens	1-1	1-2	1-3	1-4	2-2	2-3	2-4	3-3	3-4	4-4	Couples efficaces
1-1	0,6	0,4	0	0	0,4	1	0,2	1,8	2,2	4,2	1434
1-2		0,6	0,4	0,2	0,8	3,6	1,8	2,4	9,4	5	936
1-3			1,6	3,2	0,2	3	9,2	0,6	9,6	1	590
1-4				4,8	0	0,4	6,6	0	1,2	0	357
2-2					1	2,8	0,8	4	7,6	10,4	710
2-3						6,8	9,4	10,8	32	25	472
2-4							14,4	2,4	35	6,4	291
3-3								8	35	34	358
3-4									70	80	235
4-4										79	170
Couples	1437	936	590	357	710	472	291	358	235	170	

Interprétons les valeurs correspondant à la commune de Nanterre. Celle-ci produit 20 renchaînements d'alliance de type 1-1/1-1 pour 766 couples efficaces, elle en produirait donc $(20 / 766^2) / 2 = 0,68$ pour 100 couples efficaces. Pour la relation 4-4/4-4, on assiste à un effondrement du nombre de couples efficaces qui entraîne un pourcentage de 1700

renchaînements d'alliance de type 4-4 pour 100-100 couples. Le comportement de Saint Jean Trolimon est maintenant révélateur. Pour la même relation 4-4/4-4, 79 renchaînements d'alliance pour 100-100 couples; le rapport est de 1 à 4 pour la relation 2-2/2-2.

Comparaison des deux communes.

1) Dans un cas comme dans l'autre, les nombres réduits s'accroissent quand les liens s'allongent : multiplication par un facteur 100 à Saint Jean Trolimon, par un facteur 3000 à Nanterre. Les relations 1-3, 1-4 restent peu productrices dans les configurations dissymétriques : 1-3/1-1, 1-4/1-1, 1-4/2-2, 1-4/3-3 par exemple. Les configurations parfaitement symétriques, du type $i-j/i-j$ pour $0 < i, j < 5$, sont toujours plus fécondes que leurs équivalents non-symétriques. Nous avons choisi de ne représenter que celles-ci pour faire apparaître, en unités logarithmiques, la croissance selon la taille des liens (pour des raisons de lisibilité les valeurs réduites sont multipliées par 2).

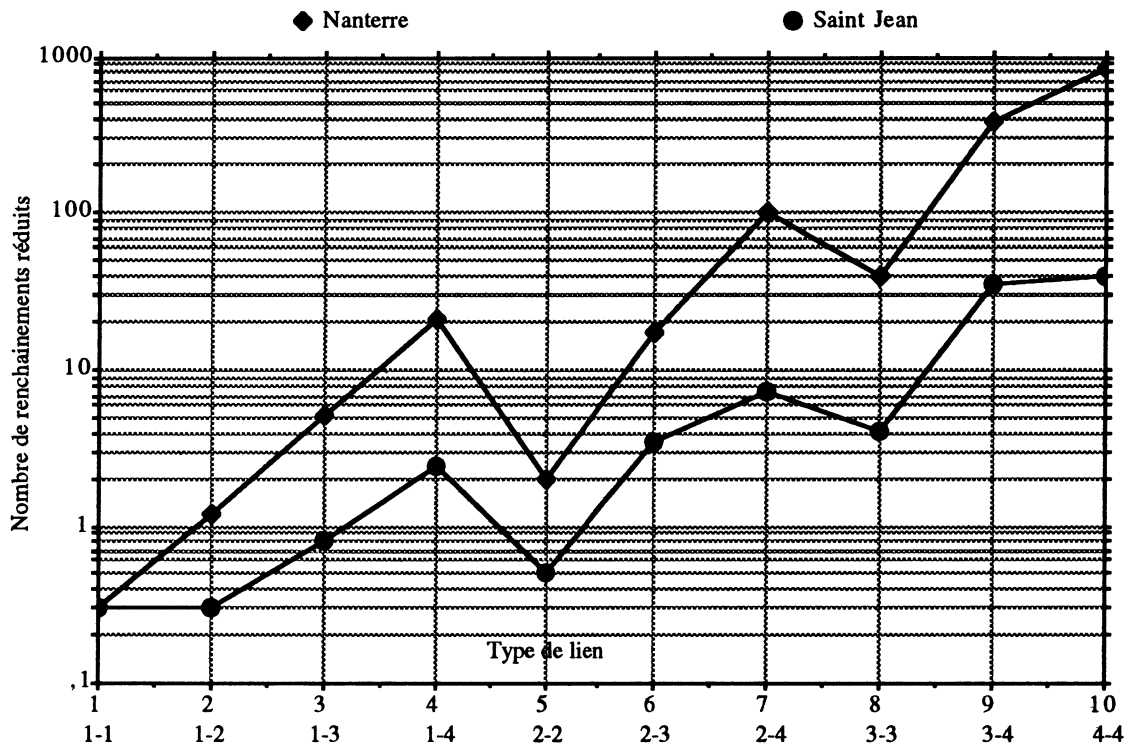


Figure 14. Représentation du nombre de renchaînements symétriques

Il n'en faudrait pas conclure pour autant que les deux communes favorisent les renchaînements correspondant à des liens distants. En effet, le nombre de possibilités pour recréer un mariage 4-4 est beaucoup plus élevé que pour un mariage 1-1. On montre facilement que le nombre de possibilités pour recréer un mariage $i-j$, dans l'hypothèse où n est le nombre d'enfants à chaque génération et où l'équilibre des sexes est respecté sur chaque fratrie est égal (égal si n impair, proche si n pair) à : $d_{ij} = \frac{(n^{i-1})(n^{j-1})}{2}$. Bien entendu, ceci n'est vrai que dans l'hypothèse, fautive partiellement, de l'indépendance des événements représentés par les mariages successifs.

En rapportant le nombre réduit aux possibilités selon les liens on fait apparaître, de façon approximative, une propension à renchaîner indépendante de la connaissance généalogique (nombre réduit) et indépendante de la profondeur généalogique des liens concernés. Nous avons choisi arbitrairement un nombre d'enfants à chaque génération de 3 pour Saint Jean

Trolimon, de 5 pour Nanterre. Un calcul moins approximatif devrait tenir compte de la distribution des fratries pour les relations 1-1, des ensembles de petits-enfants pour 2-2, etc. Quoiqu'il en soit, notre présentation fait apparaître une plus faible propension à renchaîner avec ou entre des cousins lointains. Le renchaînement entre germains est le plus prisé.

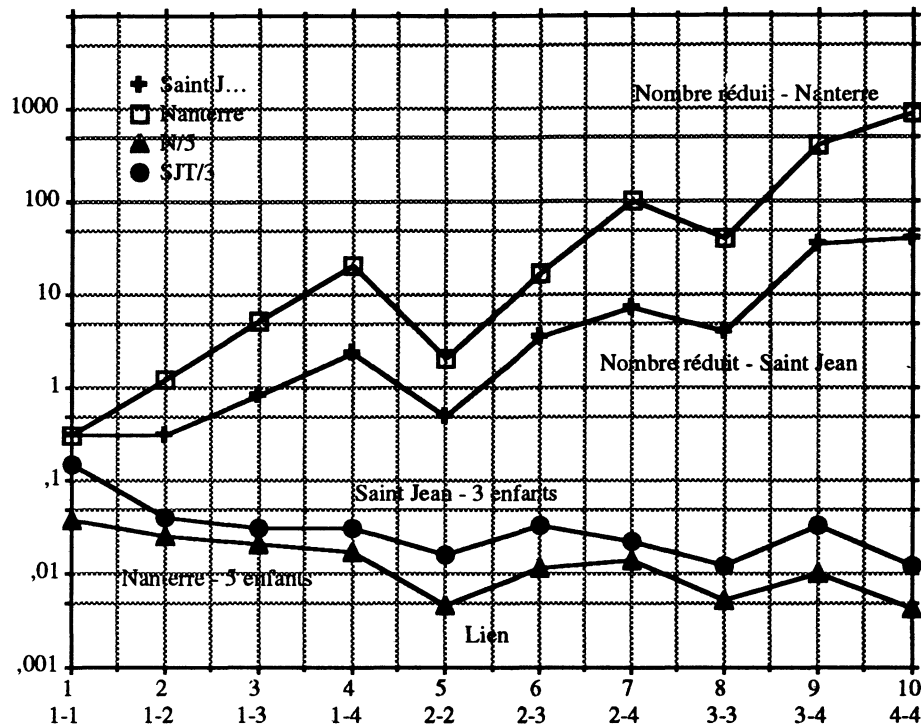


Figure 15. Représentation du nombre de renchaînements symétriques en relation avec la profondeur des liens

2) Le nombre réduit de renchaînements d'alliance est systématiquement plus élevé pour la ville de Nanterre. Les rapports selon la profondeur des liens concernés s'établissent comme suit :

1-1	1-2	1-3	1-4	2-2	2-3	2-4	3-3	3-4	4-4
1	4	6,5	9	4	5	14	10	11	12

Le calcul effectué pour passer des renchaînements bruts aux renchaînements réduits nous conforte dans l'idée que ces nombres représentent bien une différence entre les comportements des communes et non un artefact dû aux échantillons. Ainsi, le renchaînement réduit semble être une mesure efficace et fidèle de la propension à renchaîner dans une population.

3) Cette mesure fait intervenir le nombre de couples dont la connaissance de la généalogie est suffisante. Il est impossible de dire qu'elle est représentative de la population toute entière. Les renchaînements repérés sont ceux pour lesquels les généalogies sont connues. S'il existe un sous-ensemble de la population associé à une propriété et dont les généalogies sont systématiquement bien connues, nos coefficients de renchaînements réduits seront valables pour le sous-ensemble et non pour la population toute entière. Dans le cas de la ville de Nanterre, par exemple, il serait erroné de dire que cette commune produirait 850×44 , soit 37.400 renchaînements d'alliance de type 4-4 en rapport avec ses 4408 couples, si toutes les généalogies étaient connues. Il importe donc de vérifier si la population est homogène, ce qui n'est pas toujours facile.

6. RENCHAÎNEMENTS ET CLASSES SOCIO-PROFESSIONNELLES

Il se trouve justement que l'on a pu mettre en évidence, à Nanterre, une influence de la profession dans la multiplication des renchaînements d'alliance. Comme nous allons le voir, la population des agriculteurs-maraîchers-vignerons, ne se comporte pas de la même façon que le reste de la population. On peut ainsi, sur un exemple clair, vérifier l'intérêt des mesures choisies pour caractériser les populations quant aux renchaînements d'alliance.

6.1. Répartition des professions à Nanterre

Les professions des individus sont connues dans l'échantillon, directement pour les hommes, par l'intermédiaire du père de la fille, pour les femmes. L'étude de quelques lignées très riches en renchaînements montrait une sur-représentation des agriculteurs parmi les couples renchaînants. Nous avons donc regroupé les professions sur un petit nombre de classes en utilisant les règles suivantes :

- l'activité est le critère de regroupement indépendamment du niveau social dans une activité. On regroupe les agriculteurs et les ouvriers agricoles, les commandants de gendarmerie et les sergents.
- en cas d'hésitation, on s'appuie sur les associations de professions déjà réalisées dans les couples. Ainsi tous les métiers de la boucherie sont regroupés ensemble parce que l'on constate des mariages entre pratiquants de ces métiers.
- les vignerons, les agriculteurs et les maraîchers sont regroupés parce que le choix de l'appellation de ces professions évolue au cours des temps et parce que leurs comportements sont identiques.

Nous avons ainsi affecté à chaque individu une classe professionnelle et, dans un premier temps, observé la répartition de ces classes parmi les renchaînements. C'est l'objet du tableau ci-dessous :

	Total	%/Total	Renchaînements	%	Poids	% des poids / Total
Cultivateur	654	24,5%	357	53%	3451	70,2%
Vigneron	378	14%	120	18%	586	12%
Boucher	178	6,7%	29	4,3%	101	2%
Jardinier	72	2,7%	1	<1%	0	<1%
Bottier	53	2%	4	<1%	20	<1%
Militaire	69	2,6%	3	<1%	0	<1%
Carrier	129	4,9%	7	<1%	18	<1%
Epicier	38	1,4%	6	<1%	18	<1%
Menuisier	39	1,5%	2	<1%	13	<1%
Maçon	89	3,4%	2	<1%	2	<1%
Maréchal-ferrand	36	1,4%	7	<1%	25	<1%
Autres	918	34%	141	24%	565	11%
Total	2653	100%	679	100%	4856	100%

Nanterre - Répartition des renchaînements selon les professions

A ce stade de l'étude des professions, les classes sont encore peu agrégées¹⁷ et la classe "Autres" largement représentée : beaucoup de choix n'ont pas encore été faits. On constate que les 654 personnes de profession cultivateur, soit 24,5% interviennent dans 357 renchaînements, soit dans 53% des renchaînements. Si l'on tient compte des poids de ces cultivateurs, chacun d'eux intervenant dans 1, 2, ...*n* renchaînements, on vérifie qu'ils interviennent dans 70% des poids. Pour résumer ce tableau, on peut dire que les seules professions représentées dans les renchaînements sont les Cultivateurs, les Vignerons et, accessoirement les Bouchers-Charcutiers.

6.2. Répartition de la connaissance généalogique selon les professions

Comme nous l'avons déjà remarqué, dire que les seuls cultivateurs-vignerons renchaînent, c'est oublier que la répartition du nombre d'ancêtres selon les professions n'est pas égalitaire. Le tableau suivant montre qu'effectivement ce n'est pas le cas.

Génération	Ren- chaîneurs	Non- chaîneur	Boucher	Cultivateur	Carier	Jardinier	Journatier	Militaire	Vigneron
1	0	0	0	0	0	0	0	0	0
2	2	0,5	0	0	0,8	0	0	0,7	1
3	2,8	1,2	2	2,3	2	0,5	1,5	0,9	1,8
4	5,3	2,2	3,5	5,1	2,2	0,9	1,4	1,6	3,9
5	10,1	2,7	4,8	7,6	3,5	1,6	2,5	1,6	7,7
6	16,6	3,7	6,3	13,2	3,7	2,4	3,5	0,8	5,2
7	25,8	3	6,3	17,7	2,7	1,5	2,8	2,8	0
Toutes générations	12	1,9	4,3	9,6	2,9	1,4	2,8	1,3	3,8

Nanterre - Répartition du nombre d'ancêtres selon les professions

On vérifie que les renchaîneurs ont 12 ancêtres en moyenne, les non-renchaîneurs 1,9 mais surtout que les cultivateurs, les bouchers et les vignerons ont plus ou beaucoup plus d'ancêtres que les autres. Le cas des vignerons est intéressant : c'est le seul cas où la génération 5 est la mieux pourvue en ancêtres, montrant ainsi que, soit l'appellation "Vigneron" inscrite sur les actes de mariage est ancienne, soit l'activité elle-même. Ces différences dans la connaissance généalogique nous incitent à étudier, non pas la répartition du nombre brut de renchaînements selon les professions, mais bien les nombres réduits.

6.3. Répartition du nombre réduit de renchaînements selon les professions

La construction des classes a été raffinée (reclassement des "Autres") et simplifiée (agrégation de plusieurs classes proches). Nous avons repéré, pour chaque ancêtre d'une lignée renchaîneuse, le nombre de descendants de chaque classe professionnelle, le nombre total de descendants et le nombre de renchaînements. L'étude met maintenant en jeu trois classes : les métiers de la terre (paysans, maraîchers, vignerons, ouvriers agricoles,...) , les autres professions connues, les non-déclarés. Le nombre de membres de chaque profession dans

¹⁷ Cette classification est assez aventureuse. En premier lieu, nous ignorons si les comportements matrimoniaux sont liés à la relation à un patrimoine ou à une activité. Ainsi les "propriétaires" se comportent-ils comme des paysans possédant une terre ou non? Si l'on peut supposer que les professions entraînant une mobilité, les militaires par exemple, sont peu aptes à renchaîner, qu'en est-il des artisans? Il semble que les métiers de la viande, bouchers, charcutiers, transporteurs, commerçants, sont des renchaîneurs honorables : ceci peut s'expliquer par leurs relations privilégiées avec les paysans. Mais, pourquoi les maçons font-ils un score non nul?

chaque lignée est repéré. On constate que le poids brut des lignées¹⁸ (à ne pas confondre avec le poids brut des ancêtres) est une fonction parabolique du nombre de paysans dans la lignée. Le nuage est peu dispersé mais fait apparaître deux ensembles de lignées : fortement renchaîneuse (PAG2), faiblement renchaîneuse (PAG1). L'étude des poids réduits (poids brut / taille de la lignée) de ces deux ensembles est montrée ci-dessous.

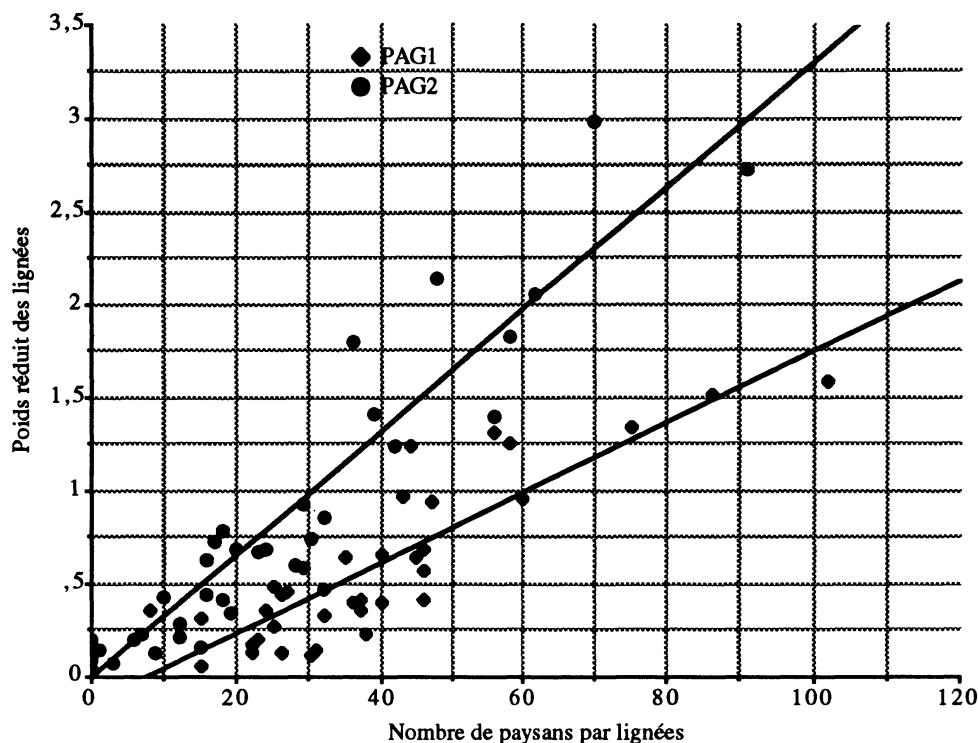


Figure 16. Représentation du poids réduit des lignées en relation avec le nombre de paysans dans les lignées

Les coefficients de corrélation sont respectivement de 0,86 et 0,95 pour les deux ensembles. On peut donc dire que le pourcentage de renchaînements par rapport à la taille de la lignée croît légèrement avec le nombre de paysans dans la lignée (pente de 0,019 et 0,033). L'étude des lignées fait apparaître une sur-représentation des lignées à fort pourcentage de non-paysans dans l'ensemble des petits renchaîneurs. Nous avons donc été amené à étudier l'influence du pourcentage de non-paysans sur le poids réduit des lignées.

¹⁸ Les différences sont généralement minimales. Elles ne deviennent importantes que dans les cas où la lignée compte quelques individus, eux-mêmes gros renchaîneurs, qui ne sont pas comptabilisés dans le poids brut des ancêtres.

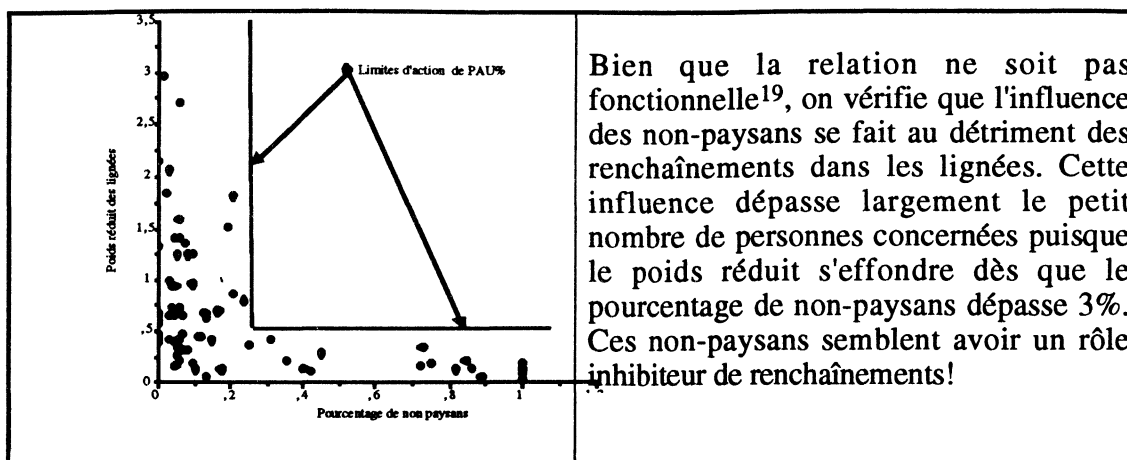


Figure 17. Représentation du poids réduit des lignées en relation avec le pourcentage de non-paysans dans les lignées

L'étude de l'influence des groupes socio-professionnels pourrait être précisée et poursuivie, en particulier en prenant en compte les bouchers-charcutiers qui jouent un rôle intermédiaire entre les non-paysans et les paysans. Il faudrait aussi contrôler les professions inconnues dont on ignore si elles sont réparties aléatoirement parmi les professions connues. Mais, dès maintenant, le constat est clair. Le renchaînement est une affaire de paysans et seulement de paysans.

7. ÉTUDE DU GRAPHE DES COUPLES RENCHAÎNANT DES ALLIANCES

Afin de montrer une intéressante voie de recherche sur les renchaînements d'alliance, nous esquissons une étude de la structure de l'ensemble des couples liés par un lien de renchaînement. Cette étude devra être poursuivie en tenant compte, en particulier, d'un ordre entre les couples.

7.1. Construction du graphe des couples renchaînants

Soit le graphe formé de l'ensemble des couples $\{A_1, A_2, \dots, A_n\}$ renchaînant au moins une fois. Les arêtes du graphe relient deux sommets $A_i A_j$ s'il existe un renchaînement entre les couples A_i et A_j . Différentes propriétés de ce graphe peuvent caractériser les populations qui nous intéressent :

- La distribution du nombre d'arêtes par sommet du graphe. Nous avons déjà étudié ce problème et nous savons qu'il existe peu de sommets munis de nombreuses arêtes et beaucoup de sommets ayant peu d'arêtes.
- Le graphe est-il décomposé en sous-graphes ou peut-on, de tout couple, atteindre tout couple, de renchaînement en renchaînement ?
- Compte tenu de la distribution hyperbolique du nombre d'arêtes, le graphe est-il homogène ou "grumeleux"? Nous reconnaissons un graphe "grumeleux" au fait que les couples lourds (à nombre d'arêtes élevé) s'associent plutôt entre eux, les couples légers faisant de même. Au contraire dans un graphe homogène, les n arêtes partant de A joignent des sommets dont le poids suit une distribution indépendante de A .
- L'effacement de certains sommets entraîne-t-il une décomposition du graphe ? Si oui, existe-t-il des sommets particuliers par lesquels on doit passer pour aller d'un couple à un autre ?

¹⁹ La transformation $x \Rightarrow 1/x$ ne permet pas de construire une régression linéaire avec un coefficient de corrélation élevée. Nous en resterons prudemment à une indication qualitative.

7.2. Propriété du graphe des couples renchaînants

Il est possible de répondre à certaines des questions précédentes. A Nanterre comme à Saint Jean Trolimon le graphe se décompose en un sous-graphe qui rassemble l'essentiel des couples et quelques sous-graphes de 2 ou 3 couples.

Taille des sous-graphes	Nanterre	Saint Jean Trolimon
2 couples	20	35
3 couples	3	5
4 couples	0	1
7 couples	0	1
306 couples / 563 couples	1	1
Total	355 couples	659 couples

Ainsi, 86 % des couples, aussi bien à Nanterre qu'à Saint Jean Trolimon forment une structure unique, les 14% restant correspondant pour l'essentiel à des paires de couples. Nous avons testé la résistance à l'émiettement de ces structures en effaçant successivement 1 puis deux sommets comptant le maximum d'arêtes. Le sous-graphe ne s'est pas séparé en fragments.

Enfin, nous avons voulu étudier la répartition des sommets à un, deux, trois, ... liens. On constate, dans les deux communes, que les sommets à un seul lien sont liés à des sommets de poids variés (de 2 à 33 liens) d'une façon tout à fait régulière. Ceci montre que, entre les deux schémas suivants, c'est le second qui représente la réalité.

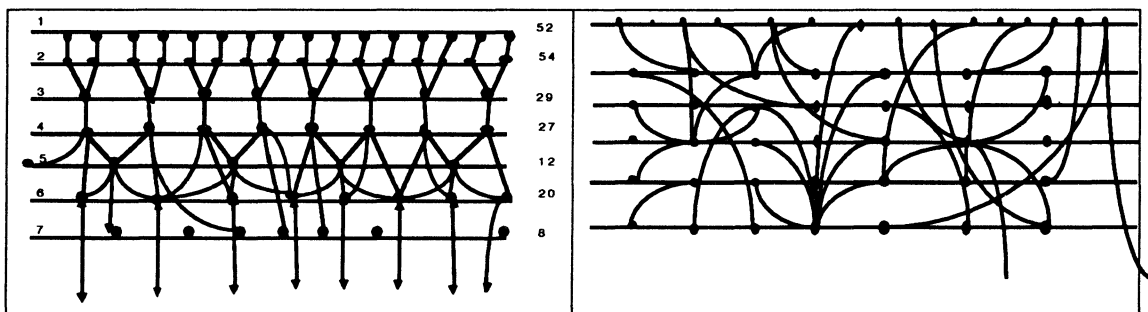


Figure 18. Représentation de la structure formée par les couples renchaînants

Les sommets les moins lourds sont reliés à tous les types de sommets et non à des sommets un peu plus lourds qu'eux, de façon progressive. Le réseau des couples n'est pas formé de couches successives de plus en plus reliées entre elles, mais d'un ensemble dans lequel tous les types de liens sont représentés (sommets lourds avec sommets lourds, sommets lourds avec sommets légers).

L'intérêt ethnographique de ces remarques n'est pas évident. D'une part, il est intéressant d'étudier les caractéristiques des sous-graphes isolés, du point de vue de leurs propriétés généalogiques et socio-professionnelles. En l'absence d'exemples plus nombreux, il est difficile de tirer des conclusions sur l'égalité du nombre de couples formant les petits sous-graphes, dans les deux communes. Il faudrait pouvoir donner un sens à cet isolement. Par contre, la nature du graphe principal est très intéressante. Elle nous montre que, bien qu'il y ait des petits et des gros renchaîneurs, il n'y a pas de classes de petits et de gros, progressivement reliés les uns aux autres. Pour une fois, riches et pauvres s'associent librement en un réseau inextricable.

Cette étude des structures serait enrichie par l'introduction d'un ordre sur les couples qui permettrait de suivre le flux des générations, des ancêtres fondateurs jusqu'aux derniers couples connus. L'ordre pourrait être fondé sur les dates au mariage. En y associant des informations de revenus, extraites des inventaires après décès, par exemple, on verrait, peut-être, clairement se construire les patrimoines.

8. CONCLUSION

Toute limitée qu'elle soit, cette étude a montré qu'il était possible de dépasser les insuffisances de l'information généalogique et de construire des mesures permettant la comparaison. Les lois de Pareto-Levy se sont révélées efficaces pour traiter les phénomènes en jeu. L'esquisse d'une analyse de la structure des couples renchaînants a montré son intérêt et ses limites. Elle doit être poursuivie et développée. L'étude de la répartition des renchaînements²⁰ d'alliance selon la taille des fratries n'a pas été présentée ici. Elle permet d'approfondir les questions liées à la gestion par les populations de ce type de mariage et, en particulier, la gestion de l'appauvrissement de la parenté affine. Les conséquences sur la transmission du patrimoine et son émiettement ne sont pas élucidées. L'importance de ce problème exige qu'il soit abordé rapidement.

BIBLIOGRAPHIE

- BARBUT M., "Des bons et des moins bons usages des distributions parétiennes en analyse des données", *Histoire et Mesure*, III-1, Paris, CNRS, (1988), 111-128.
- BARBUT M., "Distributions de type parétien et représentation des inégalités", *Math. et Sci. hum.*, n°106, (1989), 53-69.
- BARBUT M., "Note sur quelques indicateurs globaux de l'inégalité : C. Gini, V. Pareto, P. Levy", *Revue française de sociologie*, XXV, 1984.
- EYTAN M. : "Quelques propriétés de la loi de Pareto et leurs incidences en Sciences Humaines", *Math. et Sci. hum.*, n°8, (1964), 21-26.
- LEVY P., *Théorie de l'addition des variables aléatoires*, Paris, Gauthier-Villars, 1937.
- PETRUSCEWYCZ M., "Histoire de la loi d'Estoup-Zipf : documents", *Math. et Sci. hum.*, n°44, (1973), 41-56.
- RICHARD P., SEGALÉN M., "Analyse statistique des renchaînements d'alliance", in *Quinze générations de bas-bretons*, Paris, *P.U.F.*, (1985), 379-391.
- SEGALÉN M., "Quinze générations de bas-bretons", Paris, *P.U.F.*, 1985.
- SEGALÉN M., RICHARD P., "Marrying Kinsmen in Pays Bigouden Sud, Brittany", *Journal of Family History*, Vol. 11, n°2, (1986), 109-130.
- SEGALÉN M., RICHARD P., "Un isolat aux portes de Paris - Mariage et société à Nanterre entre 1750 et 1850", in *Mesurer et comprendre - Mélanges en l'honneur de Jacques Dupâquier*, Paris, *P.U.F.*, 1993.

²⁰ Comme nous l'avons déjà indiqué, les fratries sont, vraisemblablement, connues de façon incomplète à Saint Jean Trolimon. Mais, à Nanterre, on voit clairement apparaître une corrélation positive entre nombre de germains et nombre de renchaînements, concernant la fratrie. De la même façon, on constate que sur l'ensemble des germains, il en reste toujours qui ne sont ni initiateurs, ni récepteurs de renchaînements.