



HAL
open science

La parenté radiographiée. Un nouveau logiciel pour l'analyse des réseaux matrimoniaux

Klaus Hamberger, Michael Houseman, Cyril Grange

► **To cite this version:**

Klaus Hamberger, Michael Houseman, Cyril Grange. La parenté radiographiée. Un nouveau logiciel pour l'analyse des réseaux matrimoniaux. *L'Homme - Revue française d'anthropologie*, 2009, 191, pp.107-138. halshs-00445418

HAL Id: halshs-00445418

<https://shs.hal.science/halshs-00445418v1>

Submitted on 8 Jan 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

La parenté radiographiée : un nouveau logiciel pour l'analyse des réseaux matrimoniaux

Klaus Hamberger*, Michael Houseman** et Cyril Grange***
(paru dans L'Homme 191 : 107-138, 2009)

Introduction

Un paradoxe est au cœur des études de parenté : les choix matrimoniaux, qui en tant que fondement des systèmes de parenté devraient en constituer l'objet premier, ont été singulièrement négligés dans leur dimension empirique. L'anthropologie de la parenté a d'abord consisté en une analyse des terminologies et s'est ensuite développée en une science des règles et des normes ; elle se présente aujourd'hui comme un champ d'étude multiforme des représentations et des institutions, des stratégies politiques et des opérations symboliques. Mais s'il reste un aspect qui continue à y occuper une place marginale, ce sont justement les pratiques réelles qui engendrent les réseaux matrimoniaux.

La raison profonde de cette situation n'est pas tant un manque de données qu'une insuffisance de moyens et de méthodes d'exploitation. Cela tient en grande partie à la nature même de l'objet : les relations de parenté ne sauraient être traitées comme des éléments isolés. Formant système, elles n'acquièrent leur signification que par la manière dont elles se distribuent, se composent et se combinent pour former des agencements réticulaires complexes qui résistent à un traitement de type « manuel ». Il ne peut dès lors y avoir de théorie réaliste des pratiques matrimoniales sans recours à un traitement informatique.

L'ordinateur a été introduit il y a une quarantaine d'années dans les études de la parenté (Coult et Randolph 1965). Depuis les recherches innovatrices de F. Héritier (1974, 1981 ; Selz 1994) sur les bouclages consanguins et de M. Segalen (1985 ; Richard 1993) sur les renchaînements d'alliance (*cf.* Jolas *et al.* 1970), un nombre croissant de travaux témoignent de l'intérêt des outils informatiques dans ce champ d'étude, en anthropologie (Cazès et Guignard 1991, Collard 1994, Houseman et White 1996, Brudner et White 1997, Barry 1998, Ferchiou 2000, White et Johanson 2004, Sinibaldi 2007) mais aussi en histoire (Delille 1985, Grange 2005, Gasperoni 2008). Centrés pour la plupart sur l'analyse des réseaux matrimoniaux, ces travaux ont été marqués par des avancées sur le plan des formalismes (par exemple White et Jorion 1992) et par l'introduction de concepts nouveaux, telles que « structure à côtés » ou *sidedness* (Houseman et White 1996), « endogamie structurale » (White 1997), « modes de composition » (Barry 1998) ou « anneau matrimonial » (Hamberger, Houseman, Daillant, White et Barry 2004). Toutefois, ils se sont trouvés confrontés à un obstacle majeur. La plupart des logiciels disponibles ne permettaient que de rechercher certains types de mariages spécifiés au préalable, et malgré les avancées introduites par des logiciels comme Genos de Laurent Barry (1996-1997, *cf.* Barry 1996, 2004) ou Par-Calc de Douglas White (1992-1996, *cf.* White et Jorion 1992), on manquait encore d'un instrument permettant un recensement matrimonial complet d'un réseau de parenté. Or, c'est à cette seule condition qu'on peut s'attaquer à l'une des questions les plus importantes de la théorie de la parenté : celle du lien entre les choix de mariage particuliers et

* CEMAF (UMR 8171, CNRS), 27 rue Paul Bert, 94204 Ivry, tél. 01-49-60-40-11, fax. 01-46-71-84-94, klaus_hamberger@yahoo.fr.

** ECOLE PRATIQUE DES HAUTES ETUDES / CEMAF (UMR 8171, CNRS), 27 rue Paul Bert, 94204 Ivry, houseman@attglobal.net

*** CENTRE ROLAND MOUSNIER (UMR 8596, CNRS), Université de Paris IV, 1, rue Victor Cousin, 75005 Paris, cyril.grange@numericable.com

la structure du réseau global qui en émerge.

Cet article présente le logiciel Puckⁱ, qui remédie à ce manque, en ce qu'il recense la totalité des configurations matrimoniales qui composent un réseau de parenté, sans autre spécification préalable que la taille maximale des chaînes recherchéesⁱⁱ.

On se bornera ici à présenter les principales applications de ce logiciel, en précisant les enjeux théoriques et les choix méthodologiques qui ont présidé à son élaboration. En effet, la question de savoir comment représenter un réseau de parenté afin de formaliser et d'automatiser son analyse n'est pas seulement d'ordre « technique ». De même qu'une réflexion sur les propriétés des relations de parenté, loin d'être un exercice abstrait, a un impact immédiat sur la production d'outils efficaces pour la recherche empirique, inversement, les difficultés rencontrées dans l'élaboration de tels outils soulèvent des interrogations fondamentales quant à la nature même de l'objet auquel ils s'appliquent. Ecrire un logiciel comme Puck, c'est vouloir expliquer la parenté à une machine. Et comme une machine ne comprend qu'un langage sans ambiguïtés, aussi direct et aussi simple que possible, cette tâche nous mène à aborder, voire à découvrir, des problèmes théoriques élémentaires.

La problématique

Afin de faire ressortir certaines difficultés, souvent passées sous silence, que pose l'étude systématique des pratiques matrimoniales, considérons brièvement ce que représente la toute première étape d'une telle étude : une énumération des liens de consanguinité et d'affinité qui relient les conjoints au sein d'un corpus généalogique. La représentation d'une structure matrimoniale empirique (que ce soit en mots, en images ou en chiffres) ne se réduit certes pas au seul recensement de ces liens. Cependant, pour comprendre comment ces relations se distribuent, la façon dont elles s'assemblent pour former des chaînes, des circuits et des réseaux, il faut d'abord savoir quelles sont-elles et combien il en existe.

De nombreux ouvrages font état des fréquences de tel ou tel type de mariage dans une population donnée. Un cas souvent rencontré, quoique quelque peu caricatural, est celui d'un tableau rapportant le pourcentage des unions entre cousins germains. Constatant, par exemple, que les fréquences des mariages avec FBD, MBD, FZD et MZDⁱⁱⁱ sont en ordre décroissant, l'auteur pourrait conclure que la société en question affiche une préférence pour les unions de type FBD, un évitement des mariages MZD, et une indifférence relative pour les autres types de cousins. Or une telle conclusion repose, tacitement, sur certaines hypothèses.

Tout d'abord, on suppose (1) que les unions entre cousins germains (en plus sans différenciation entre pleins et demi-germains) sont, à elles seules, suffisantes pour définir le système de préférences et d'évitements matrimoniaux en vigueur, sans qu'il soit nécessaire de connaître, par exemple, les fréquences des mariages entre cousins issus de germains ou entre affins. Ensuite, on estime que (2) le corpus utilisé est représentatif des choix matrimoniaux de l'ensemble de la population que l'on cherche à étudier, soit que ce corpus constitue un ensemble d'individus reliés les uns aux autres de façon suffisamment dense et homogène pour que leurs choix de mariage soient régis par un même système d'évitements et de préférences. Enfin, il est présumé (3) que les effectifs de mariage *recensés* entre cousins donnent, en soi, un juste reflet de leurs fréquences *réelles*.

Or, rien ne permet de partir *a priori* de ces hypothèses. Il suffit, par exemple, de trouver que la plupart des MBD épousées sont en même temps des FFBSD (*cf.* Barry 1998), où bien qu'elles sont des BW (en raison d'un régime léviratique), pour que les fréquences de ces mariages perdent leur valeur d'indicateur d'une préférence matrimoniale indépendante (hypothèse 1) ; il suffit que tous les mariages FBD soient localisés dans une partie du corpus,

et tous les mariages MBD dans une autre, et que ces deux ensembles ne soient connectés que par un petit nombre de chaînes de parenté longues, pour que les chiffres établis cessent de représenter le profil matrimonial d'une seule et même population (hypothèse 2) ; enfin, il suffit de trouver que le corpus est biaisé du côté agnatique (par exemple en raison d'un régime de résidence patrilocal) et que le nombre de cousines agnatiques disponibles dépasse largement celui des cousines utérines, pour que l'excédent des mariages recensés de type FBD ne soit plus, à lui seul, un bon indicateur d'un excédent réel (hypothèse 3).

Le caractère tendancieux des énumérations des types de mariage rapportées dans bon nombre de travaux est toutefois moins gênant qu'on pourrait le penser, en raison de la place périphérique que leur assigne l'analyse. Le plus souvent, ces tableaux statistiques ne servent pas à tirer des conclusions, mais à confirmer et à illustrer des conclusions auxquelles le chercheur est parvenu par d'autres moyens, comme, par exemple, l'analyse de la terminologie, ou des prohibitions et prescriptions matrimoniales. L'étude des pratiques matrimoniales devient ainsi un appendice et un supplément à l'analyse des schèmes de classification, des normes et des discours. Mais ce sont justement ces schèmes, normes et discours qui biaisent souvent, sinon le regard du chercheur lui-même, tout au moins les énoncés et la mémoire de ses informateurs, de sorte que la soi-disant « confirmation » du modèle par les pratiques n'est parfois que celle d'une prédiction dont la validation est assurée d'avance. Au lieu de servir de contrôle et d'instrument critique pour des modèles théoriques (y compris les modèles « émiques »), l'analyse des régularités empiriques s'avère ainsi, dans le meilleur des cas, un exercice superflu, au pire, un sophisme par lequel on tire du chapeau le lapin qu'on vient d'y mettre.

Un recensement pertinent des liens de parenté entre conjoints doit répondre à certaines conditions qui sont difficiles à remplir sans utiliser des moyens informatiques. D'abord et surtout, il doit être exhaustif (dans certaines limites spécifiées). Ensuite, on doit pouvoir le restreindre à des sous-ensembles à même de servir de bases adaptées aux objectifs de la recherche. Enfin, il faut qu'il soit possible d'identifier et de quantifier les biais du corpus utilisé, et de corriger, si possible, les chiffres bruts.

Le logiciel Puck a été développé pour répondre à ces exigences. Il dénombre l'*ensemble* des relations de parenté entre conjoints dans les limites spécifiées par leur taille maximale^{iv}, et sans spécification préalable des types de mariage recherchés. Il rend possible la segmentation des corpus selon des critères endogènes ou exogènes à la parenté pour extraire des sous-corpus plus restreints, et des recensements matrimoniaux globaux ou ciblés sur ces sous-corpus. Enfin, il permet d'évaluer, par plusieurs indicateurs, la qualité du corpus, d'identifier ses divers biais, et de produire, par une simulation contrôlée, des structures aléatoires pour pouvoir saisir les divergences entre des distributions réelles et simulées des fréquences de mariages.

Un logiciel au croisement de deux approches

Le concept d'une relation de parenté, et plus généralement d'une relation sociale, peut se formuler de deux manières : soit de façon « ego-centrée », soit de façon « socio-centrée ». Dans une perspective ego-centrée, une relation sociale fait partie, en quelque sorte, des propriétés d'un individu : Ego a un père, une mère, un conjoint, etc., tout comme il a un âge, un sexe ou une profession. En revanche, dans une perspective socio-centrée, une relation est un objet en soi, qui existe au même titre que les individus. Dans la première perspective, l'univers social se présente comme un ensemble d'individus dont chacun reflète, comme en miroir, son environnement personnel. Dans la seconde, il est représenté par un ensemble d'individus et par l'ensemble des chemins qui les connectent – autrement dit, par un réseau.

La divergence entre ces deux approches s'impose lorsque l'on s'attache à construire une relation de parenté complexe, comme, par exemple la relation « frère », composé de la relation « père » ou « mère » et de la relation « fils ». D'un point de vue ego-centré, cette construction se comprend comme une série de transformations qui met successivement Ego à la place d'Alter : un « frère » est l'individu qui devient mon « fils » si je me mets à la place de mon « père » ou de ma « mère ». Repérer les relations de type « frère » dans un corpus généalogique consiste donc à identifier, pour chaque individu, l'ensemble de ses parents, et, à déterminer, pour chaque parent d'un individu donné, l'ensemble de ses fils. La construction de la relation procède en parcourant un *arbre*.

D'un point de vue socio-centré, construire la relation « frère » revient à composer cette relation à partir d'autres relations plus simples : un chemin du type « frère » résulte d'une combinaison d'un chemin du type « père » ou « mère » avec un chemin du type « fils ». On regardera donc, pour chaque paire d'individus, s'il existe un individu tiers qui est « père » ou « mère » du premier et qui a le deuxième pour « fils ». La construction est fondée sur la manipulation d'une *matrice*.

Ces deux méthodes se distinguent à la fois dans leurs exigences techniques (en temps de recherche et en espace de mémoire) et dans leurs implications épistémologiques. Dans une optique égocentrée, une chaîne de parenté n'existe que du point de vue particulier d'un Ego donné ; chacun des liens qui la composent doit être repéré de nouveau en partant de chaque individu, même s'il a déjà été maintes fois parcouru en partant d'autres individus. Dans un réseau dense, cette méthode peut prendre un temps considérable. Sous un angle sociocentré, décrire une relation de parenté consiste à indiquer, pour toute paire d'individus possible, s'ils sont ou non reliés par elle. Dans un réseau large, cette méthode peut exiger un grand espace de mémoire.

La différence des deux approches ne se réduit pas à ces aspects techniques. Il s'agit en effet de la conception d'une chaîne de parenté en tant que telle, question d'autant plus importante si cette chaîne est close par un mariage. Prenons, par exemple, le cas d'un homme qui épouse la sœur de la femme de son frère (BWZ). Cette situation peut être décrite de quatre façons différentes, selon le point de vue de chacun des quatre individus impliqués. Dans une optique égocentrée, ces quatre formules représentent des chaînes tout à fait différentes qui impliquent des choix et des normes sociales distinctes, et doivent en conséquence être considérées séparément lors d'un recensement matrimonial. D'un point de vue sociocentré, ces différentes formules ne constituent que quatre manières alternatives de représenter une seule et même configuration matrimoniale, qui ne doit être comptée qu'une seule fois pour éviter des doubles comptages.

Toute modélisation d'une structure de parenté doit alors décider jusqu'à quel point elle met en œuvre un engendrement des relations de parenté par un processus de substitution d'Alter par Ego (approche ego-centrée), et à quel moment elle reconnaît aux relations une objectivité au-delà des points de vue individuels impliqués (approche socio-centrée). Le logiciel Puck se veut une synthèse de ces deux approches.

Le développement de Puck s'inscrit en effet dans une double généalogie. D'une part, il poursuit, bien que d'une façon sensiblement différente, la démarche initiée par Laurent Barry avec le logiciel Genos (Barry 1996, 2004). Puck ne reprend ni le langage de programmation (Microsoft Visual Basic), ni les algorithmes de Genos^v, mais est à plusieurs égards l'héritier spirituel de ce logiciel. En particulier, Puck retient la notation novatrice introduite par L. Barry (voir la section suivante).

D'autre part, Puck s'inscrit dans la continuité des travaux menés sur les réseaux de parenté à partir des concepts de la théorie des graphes. Il s'agit en premier lieu du logiciel P-graph^{vi} développé (en Fortran) par Douglas White (White et Jorion 1992 ; Houseman et White 1996, 1998), puis plus récemment, de l'application aux réseaux de parenté du logiciel Pajek^{vii},

logiciel qui permet une analyse rapide des grands réseaux comme le sont fréquemment les corpus généalogiques (cf. White, Batagelj et Mrvar 1999, Hamberger et al. 2004, Sinibaldi 2007, Batagelj et Mrvar 2008)^{viii}.

Les deux logiciels évoqués – Genos et Pajek – se distinguent par la mise en œuvre presque exclusive d'un des deux approches complémentaires, égocentrée et sociocentrée, qu'on vient d'esquisser. Genos est le fruit d'une conception essentiellement ego-centrée de la parenté, selon laquelle aucune chaîne de parenté ne peut être permutée sans changer sa nature. En revanche, Pajek adopte une approche foncièrement socio-centrée qui représente toute relation sociale comme un réseau. Puck utilise une approche synthétique qui combine les vertus de ces deux méthodes.

Concepts et méthodes

La notation positionnelle

Les calculs centraux de Puck, ainsi que certains éléments de présentation, se fondent sur les principes de la notation positionnelle des relations de parenté développée par L. Barry (2004). Cette notation représente les relations de parenté par des chaînes composées de trois lettres désignant le sexe des individus – « H » (homme), « F » (femme), « X » (individu de sexe inconnu) – et de deux signes diacritiques : le point « . » pour indiquer un lien de mariage, et des parenthèses « () » pour indiquer le ou les ascendants apicaux dans la chaîne^{ix}. Ces chaînes se lisent de gauche (Ego) à droite (Alter) ; les liens entre les H, F et X successifs sont ascendants à gauche et descendants à droite des parenthèses apicales. Par exemple, la fille du frère de la mère, qui en notation standard s'écrit MBD, implicitement pour un Ego masculin, est ici HF()HF (les parenthèses vides sont équivalentes à « (HF) », c'est-à-dire à des individus apicaux des deux sexes). Le sexe d'Ego est rendu explicite – si Ego était féminin, la chaîne s'écrit FF()HF – ainsi que la relation de pleine germanité entre la mère et son frère – si c'était la fille du frère agnatique de la mère, la chaîne, intégrant donc un seul ancêtre apical, s'écrit HF(H)HF. La transparence de cette écriture, qui fournit une image immédiate de l'architecture des relations de parenté, la rend supérieure à la notation standard fondée sur des abréviations des termes anglais, non seulement pour la programmation mais aussi comme instrument d'analyse et de réflexion théorique.

Le concept de circuit matrimonial

Le concept au fondement des algorithmes de Puck est celui de « circuit matrimonial ». Un circuit matrimonial est toute chaîne dans un réseau généalogique qui est fermée sur elle-même et au sein de laquelle un lien de descendance ne peut être immédiatement suivi par un lien d'ascendance. Cette condition implique qu'elle passe nécessairement par au moins un lien de mariage (d'où son nom). Un circuit matrimonial peut être interprété comme une chaîne de parenté reliant des conjoints. La figure 1a fournit l'exemple d'un circuit simple comportant trois mariages. La nature exacte des liens d'ascendance (asc.1, 2, 3) et de descendance (desc.1, 2, 3) n'est pas représentée ; les individus 23 et 33 y sont situés respectivement comme Ego et Alter. Mais notons qu'on peut tout aussi bien identifier ce même circuit par une autre chaîne qui situe un autre couple comme Ego et Alter, les individus 37 et 64 par exemple (Figure 1b). En fait, d'un point de vue socio-centré, l'ensemble des chaînes qui contiennent les mêmes liens dans le même ordre sont des représentations alternatives d'un seul et même circuit matrimonial.

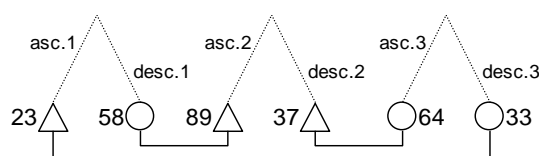


Figure 1a

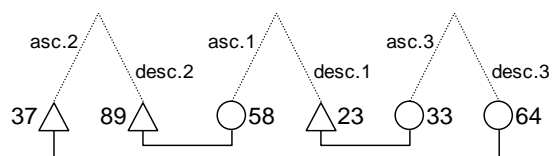


Figure 1b

Les liens de consanguinité et d'affinité qui constituent les circuits matrimoniaux n'existent pas de façon isolée, mais s'associent les uns aux autres pour former des systèmes réticulaires : chaque connexion entre deux individus fait partie de plusieurs, voire d'un nombre virtuellement infini de chaînes de parenté. Le lien entre moi et mon père, par exemple, fait partie des chaînes entre moi et le père de mon père, entre moi et la sœur de mon père, entre la sœur de mon père et mon épouse, entre mon épouse et le mari de la sœur de mon père, etc. De façon similaire, un mariage donné peut faire partie non pas d'un seul circuit matrimonial mais de plusieurs. Le mariage « A » dans la Figure 2a par exemple, participe à la fois au circuit 1-2-3-4-1 et au circuit 1-2-5-6-7-8-4-1^x. Plus encore, un circuit matrimonial donné peut soit entrer dans la composition d'un autre circuit, soit être complètement inclus dans un circuit plus large. Considérons la Figure 2b. Le circuit 1-3-6-7-8-1 et le circuit 2-4-5-6-3-2 composent ensemble le circuit 1-2-4-5-6-7-8-1 : l'individu 1 s'est marié avec la fille du frère de sa mère (MBD ou HF(H)HF en notation positionnelle) qui, en raison du mariage de son père, 2, avec la fille du frère de son propre père (sa FBD ou HH(H)HF), est aussi la fille du fils du frère du père de son père (FFBSD ou HHH(H)HHF). En même temps, le circuit 1-3-6-7-8-1 est entièrement inclus dans le circuit 1-2-3-6-7-8-1 qui situe 3, la mère de 1, comme l'épouse de son père 2, et situe 8 comme une affine plutôt que comme une consanguine : la fille du frère de l'épouse du père (FWBD ou H(H).F(H)HF). Dans ce cas, à la différence de la figure précédente, tous les individus faisant parti d'un circuit (1-3-6-7-8-1) font également parti de l'autre (1-2-3-6-7-8-1)^{xi}.

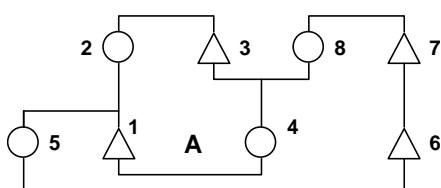


Figure 2a

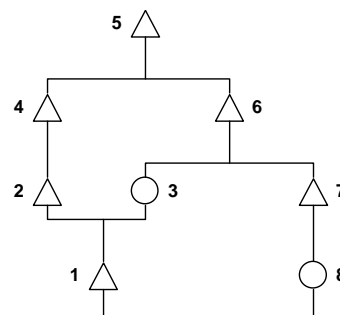


Figure 2b

Se chevauchant, s'enchaînant les uns dans les autres, les circuits matrimoniaux ne sont donc pas mutuellement indépendants, ni sur le plan formel, ni, dans bien des cas, sur le plan sociologique. Ainsi, un ensemble de mariages raccordés entre eux sont à envisager non pas comme une simple juxtaposition de comportements individuels ou collectifs, mais comme une coordination dynamique de ces comportements. Si l'on observe, par exemple, que les unions de type MBD sont plus fréquents lorsque le père du mari a épousé sa cousine parallèle patrilatérale (FBD), il se peut bien que cette corrélation s'explique par le fait que la combinaison de ces deux unions fait du mariage de type MBD en même temps un mariage de

type FFBSD (Figure 2b, *cf.* Barry, 1998). En raison de cette interdépendance des circuits matrimoniaux au sein du réseau de parenté, toute tentative de caractériser sa structure exige la prise en compte de la distribution de *tous* les types de circuits dans les limites d'un critère aussi neutre que possible (comme la taille des chaînes). Le recensement des circuits matrimoniaux d'un seul type ou même de plusieurs types présélectionnés (les unions avec les cousins germains par exemple, *cf. supra*), peut s'avérer un indicateur incomplet voire trompeur.

Le problème se complique par le fait que, comme on l'a vu, chaque circuit peut être représenté de multiples façons selon l'individu qui est pris comme Ego. Le circuit 1-5-6-7-8-4-1 auquel est intégré l'union A dans la Figure 2a peut être représenté de quatre façons : comme ZHFZD (du point de vue de l'individu 1 : 1-5-6-7-8-4-1), comme FZDHZ (du point de vue de l'individu 6 : 6-7-8-4-1-5-6), comme MBSWMS (du point de vue de l'individu 4 : 4-8-7-6-5-1-4), et comme MSWMBS (du point de vue de l'individu 5 : 1-4-8-7-6-5). Vouloir envisager ces configurations comme des types d'unions différentes plutôt que comme des représentations alternatives d'un seul et même circuit est parfaitement défendable (Puck offre aussi cette possibilité). Toutefois, afin de procéder à un recensement global des circuits matrimoniaux au sein du réseau, il est indispensable d'éviter de tels comptages multiples.

La technique de recherche de circuits

Sous de telles conditions, une recherche exhaustive des circuits matrimoniaux au sein d'un réseau de parenté doit faire intervenir deux types de procédés : une recherche des chaînes de parenté au sein du réseau, et une sélection parmi ces chaînes en fonction de leurs propriétés. Ces propriétés interviennent à deux niveaux : d'un côté pour identifier les chaînes qui ont une forme de circuit, et de l'autre, pour s'assurer qu'aucun circuit identifié ne soit compté deux fois (ce qui implique un choix entre ses multiples représentations possibles).

Si l'on adopte une perspective égocentrée, comme c'est le cas de Genos, on procède d'abord à une recherche de chaînes de parenté à partir de chaque Ego, et ce n'est que dans un deuxième temps que la prise en compte des propriétés socio-centrées des circuits intervient afin d'éliminer les chaînes mal formées ou redondantes. A l'autre extrême, entièrement socio-centré, une recherche sous Pajek prend pour objet le réseau de parenté tout entier, soit pour le soumettre à une série de transformations (multiplications de matrices), soit pour y repérer des configurations des circuits préétablies, conçues comme « fragments » de la structure globale du réseau. Comme nous l'avons déjà indiqué dans la section précédente, l'approche de Genos implique un temps de calcul important pour produire des chaînes dont une grande partie sera ensuite éliminée ; en revanche, l'approche de Pajek exige soit un important espace de mémoire, soit un immense inventaire, établi à l'avance, des formes de circuits à rechercher. La méthode de Puck consiste à combiner ces deux approches en faisant intervenir les propriétés socio-centrées des circuits dès le départ pour orienter la recherche ego-centrée des chaînes de parenté.

Cette orientation repose sur une modélisation particulière de la structure d'un circuit matrimonial. Comme nous l'avons déjà indiqué, tout circuit matrimonial peut être représenté comme une chaîne de liens de parenté où le premier et le dernier individu sont liés par un mariage. Ainsi, toute chaîne qui constitue un circuit matrimonial est fondée sur une « armature » formée par des individus mariés, reliés entre eux par une série alternée de liens de consanguinité et de mariage (Figures 1a et 1b). Chacun des individus inclus dans cette armature peut figurer comme Ego dans autant de représentations possibles du circuit qu'il a des conjoints dans ce circuit (généralement une seule, et dans certains cas, comme celui d'un homme marié à deux sœurs, deux). En conséquence, sélectionner une représentation unique pour chaque circuit afin qu'il soit compté une fois et une seule, revient à sélectionner un Ego

– et le cas échéant aussi un Alter – parmi les individus mariés de la chaîne. Ceci peut se faire au moyen d'un critère simple comme, par exemple, le numéro d'identité le moins élevé (selon ce critère, c'est la chaîne de la Figure 1a, situant l'individu 23 comme Ego, qui sera retenu). Cette sélection est indépendante des individus intermédiaires par lesquelles passent les chaînes de consanguinité. Pour pouvoir sélectionner les circuits partant d'un Ego donné, il suffit donc, dans un premier temps, de construire leurs armatures, composées de liens de mariage et de liens de consanguinité non spécifiques. Le repérage de ces liens ne nécessite pas de détailler des chaînes de consanguinité et peut donc être effectuée, une fois pour toutes, au début de la procédure de recensement, sans être répétée à nouveau pour chaque Ego. La sélection des armatures se fait au fur et à mesure de leur construction, par la comparaison des numéros de chaque conjoint ou consanguin qu'on souhaite ajouter à la chaîne avec les numéros des individus qui le précèdent dans la chaîne. On évite ainsi la construction de chaînes redondantes ou mal formées, de sorte que seules les armatures « admissibles » seront utilisées pour construire des circuits en détaillant les chaînes de consanguinité qu'ils renferment. En répétant ce processus pour chaque individu marié du réseau, tous les circuits matrimoniaux peuvent être trouvés sans qu'aucun ne soit compté deux fois. Ainsi, des techniques de construction ego-centrées et socio-centrées concourent à réduire le temps de calcul et à minimiser l'espace mémoire nécessaire.

Eviter des doubles comptages de cette façon ne signifie nullement qu'au bout du processus les différentes représentations d'un circuit ne soient plus repérables. Selon le choix de l'utilisateur, les résultats du recensement matrimonial peuvent être organisés d'une façon ego-centrée (par exemple sous la forme de listes triées par Ego), ou de façon « socio-centrée » (par exemple sous la forme de réseaux directement lisibles par Pajek). Ces résultats peuvent ainsi permettre le développement de modèles reposant sur des principes différents. Cette neutralité est même un trait voulu du logiciel, dont l'emploi doit justement fournir une base objective et commune à des modèles divergents pour pouvoir comparer leurs puissances et leurs faiblesses relatives. C'est ainsi qu'on a souhaité rendre aux pratiques matrimoniales empiriques leur « existence » propre par rapport aux propos théoriques et normatifs des chercheurs et de leurs informateurs, afin qu'elles puissent intervenir non pas à titre d'illustrations de modèles préétablis, mais comme instance de contrôle, et idéalement, comme fondement de la modélisation.

Le recensement matrimonial

Résultats sous forme de tableaux

Pour recenser globalement tous les circuits matrimoniaux d'un corpus généalogique dans les limites de certaines dimensions, il suffit de spécifier le nombre de liens de mariage qu'ils intègrent, et le degré canon maximal des chaînes de consanguinité qui les composent^{xii}. Toutefois, Puck permet aussi de restreindre les recensements à un seul type de circuit, ou aux seuls circuits ayant certains traits en commun (par exemple l'ensemble des unions avec une parente paternelle ou maternelle, ou tous les mariages avec la sœur d'un beau-frère ou d'une belle-sœur, etc.). Pour faciliter de telles recherches, Puck intègre une sorte de « calculette » qui identifie les propriétés caractéristiques d'une relation de parenté donnée^{xiii}, et sert d'interface entre l'analyse théorique des relations de parenté et leur recensement dans un corpus réel.

Les résultats d'un recensement matrimonial peuvent être affichés et sauvegardés sous différents formats. Les Tableaux 1 à 4, par exemple, montrent des extraits d'un recensement du corpus généalogique des Amérindiens Trio (Rivière 1969)^{xiv} pour les unions consanguines

(circuits de largeur 1) jusqu'au 3^{ème} degré canon (hauteur 3), les renchaînements impliquant deux liens d'affinité (largeur 2) jusqu'au 3^{ème} degré canon (hauteur 3) et les renchaînements impliquant trois liens d'affinité (largeur 3) jusqu'au 2^{ème} degré canon (hauteur 2).

On voit dans le Tableau 1 que 12 unions interviennent dans 13 circuits de mariage consanguin de 10 types différents, lesquels sont listés en notations standard et positionnelle, avec le nombre d'individus impliqués (Indiv) et le pourcentage qu'ils représentent parmi tous ceux intervenant dans un circuit consanguin (% Indiv), le nombre de mariages impliqués (Mar) et le pourcentage qu'ils représentent parmi tous ceux intervenant dans un circuit consanguin (% Mar), la fréquence des circuits en chiffres absolus (Circuits) et comme pourcentage de l'ensemble des circuits consanguins (% Circuits), ainsi que d'autres informations (selon le choix de l'utilisateur) sur les circuits concernés, en l'occurrence la longueur de la chaîne de parenté impliquée (LONG.) et la nature des composantes consanguines qu'intègre cette chaîne (LIGNE : U = utérin, A = agnatique, C = cognatique, I = individu [unions polygames], B = pleine germanité)^{xv}. 107 mariages interviennent dans 138 circuits de renchaînement avec deux liens d'affinité (circuits de largeur 2) de 88 types différents, et 190 mariages interviennent dans 847 circuits de renchaînement avec trois liens d'affinité (circuits de largeur 3) de 563 types différents.

Tableau 1. Liste globale des fréquences.

Trio [3, 3, 2, 0] Recensement de circuits matrimoniaux Mon Mar 09 15:32:59 CET 2009										
largeur maximale (liens de mariage): 3 hauteur maximale (distances générationnelles): [3, 3, 2, 0] pleins germains assimilés tous circuits considérés (pas de restriction aux anneaux)										
314 relations de mariage (242 hommes, 237 femmes)										
Vue d'ensemble:										
12 mariages (3.82%) impliquant 23 (4.73%) individus (11 hommes, 12 femmes) dans 13 circuits de 10 types différents (fréquence moyenne 1.3) pour 1 lien de mariage et une profondeur générationnelle maximale de 3										
Id	Standard	Positionnelle	Indiv	% Indiv	Mar	% Mar	Circuits	% Circuits	LONG.	LIGNE
1	ZD	H()FF	2	8.7	1	8.33	1	7.69	3	U
2	FDD	H(H)FF	2	8.7	1	8.33	1	7.69	3	C
3	MSD	H(F)HF	2	8.7	1	8.33	1	7.69	3	C
4	MDD	H(F)FF	2	8.7	1	8.33	1	7.69	3	U
5	FFDD	HH(H)FF	4	17.39	2	16.67	2	15.38	4	C
6	MBD	HF()HF	5	21.74	3	25.0	3	23.08	4	C
...										
107 mariages (34.08%) impliquant 176 (36.21%) individus (85 hommes, 91 femmes) dans 138 circuits de 88 types différents (fréquence moyenne 1.57) pour 2 liens de mariage et une profondeur générationnelle maximale de 2										
Id	Standard	Positionnelle	Indiv	% Indiv	Mar	% Mar	Circuits	% Circuits	LONG.	LINE
11	FW	H(H).F	6	3.41	4	3.74	2	1.45	2	A I
12	WM	H.F(F)	6	3.41	4	3.74	2	1.45	2	I U
13	BW	H(H).F	7	3.98	6	5.61	3	2.17	3	B I
14	MSW	H(F)H.F	3	1.7	2	1.87	1	0.72	3	U I
15	WMM	H.FF(F)	3	1.7	2	1.87	1	0.72	3	I U
16	WZ	H.F()F	15	8.52	10	9.35	5	3.62	3	I B
17	WFD	H.F(H)F	8	4.55	7	6.54	4	2.9	3	I A
...										
43	BWMD	H(H).F(F)F	4	2.27	2	1.87	1	0.72	5	B U
44	ZHFD	H()F.H(H)F	8	4.55	4	3.74	2	1.45	5	B A
45	ZHMD	H()F.H(F)F	4	2.27	2	1.87	1	0.72	5	B U
46	FWBD	H(H).F()HF	13	7.39	7	6.54	4	2.9	5	A A
47	FWMSD	H(H).F(F)HF	9	5.11	5	4.67	3	2.17	5	A C
...										
190 mariages (60.51%) impliquant 303 (62.35%) individus (153 hommes, 150 femmes) dans 847 circuits de 563 types différents (fréquence moyenne 1.5) pour 3 liens de mariage et une profondeur générationnelle maximale de 2										
Id	Standard	Positionnelle	Indiv	% Indiv	Mar	% Mar	Circuits	% Circuits	LONG.	LINE
99	WFW	H.F(H).F	8	2.64	6	3.16	2	0.24	3	I A I
100	MHFW	H(F).H(H).F	5	1.65	3	1.58	1	0.12	4	U A I
101	FWSW	H(H).(F)H.F	16	5.28	11	5.79	7	0.83	4	A U I
102	WMFW	H.FF(H).F	4	1.32	3	1.58	1	0.12	4	I C I
103	WFWM	H.F(H).F(F)	5	1.65	3	1.58	1	0.12	4	I A U
104	WFWD	H.F(H).(F)F	47	15.51	32	16.84	24	2.83	4	I A U
105	FFWHD	HH(H).F.(H)F	6	1.98	4	2.11	2	0.24	5	A I A
...										

Guide de lecture. La première ligne se lit de la façon suivante : le premier circuit recensé du corpus est le mariage avec la fille de la sœur (ZD en notation standard et H()FF en notation positionnelle [les parenthèses vides sont équivalentes à (HF)]); ce circuit implique 2 individus (Indiv) qui représentent 8.7% de ceux impliqués dans des circuits consanguins (% Indiv), et un seul mariage (Mar) qui représente 8.33% des unions consanguines ; il s'est réalisé une fois (Circuits) et représente 7.69% des circuits consanguins au sein du corpus ; il a une longueur (degré civil) de 3 (LONG.) et comporte une composante consanguine utérine (LIGNE).

Cette première liste globale des fréquences est complétée par trois autres : (1) pour chaque circuit matrimonial recensé, une liste nominative des couples concernés et des numéros d'identité des individus composant la chaîne de parenté qui les relie (Tableau 2) ; (2) pour chaque couple, les différents types de circuit matrimonial auxquels correspond leur union ainsi que les chaînes de parenté correspondantes (Tableau 3) ; enfin (3) l'ensemble des informations contenues dans les listes précédentes, reproduites sous la forme d'une liste unitaire qui peut être facilement triée (Tableau 4). Selon les options choisies, cette dernière liste peut aussi afficher chaque circuit matrimonial recensé dans toutes ses représentations possibles.

Tableau 2. Liste par types de mariage (circuits matrimoniaux)

Liste détaillée (par type de mariage)		
1. ZD (1)	Asanri – Eukiye	451 (452 540) 356 454
2. FDD (1)	Poto - Apok?nini	494 (495) 492 493
3. MSD (1)	Matato - Numepo	31 (24) 23 29
4. MDD (1)	Kapai Iputipo - Napanako	315 (316) 314 313
5. FFDD (2)	Pesaipo - Piruru	93 91 (90) 63 72
	Kamapo – Mori	83 89 (90) 124 87
...		

Guide de lecture. La première ligne se lit de la façon suivante : le circuit 1, le mariage avec la fille de la sœur (ZD), est réalisé une fois au sein du corpus ; le époux du couple est Asanri (Ego) et son épouse Eukiye (Alter) ; la chaîne de parenté qui les relie est représentée selon les conventions de la notation positionnelle : elle monte de l'individu 451 (Ego) aux individus 452 et 540 (ancêtres apicaux qui sont eux-mêmes des conjoints) puis descend à l'individu 356 (la sœur d'Ego) puis à l'individu 454 (Alter).

Tableau 3. Liste par couples

Liste détaillée (par couples)		
192 couples impliqués		
1. Susuku = Mokuri (2)		
	DHMHM	(1) 6 . 9 (12) . 7 (3)
	DHFZHMFD	(1) 6 . 9 11 (33 34) 28 . 23 24 (212) 3
2. Pesa = Mokuri (6)		
	SWSWM	(5) 7 . (12) 9 . 6 (3)
	DHDHMFD	(5) 8 . (14) 20 . 114 24 (212) 3
	DHMDHMFD	(5) 8 . 14 (15) 55 . 52 24 (212) 3
	DHFFDHMFD	(5) 8 . 14 235 (231) 55 . 52 24 (212) 3
	DHMDDHMFD	(5) 8 . 14 (15) 34 28 . 23 24 (212) 3
	DHFFDDHMFD	(5) 8 . 14 235 (231) 34 28 . 23 24 (212) 3
3. Sere = Pakori (48)		
	MHMD	9 (12) . 7 (3) 6
	FZHMFD	9 11 (33 34) 28 . 23 24 (212) 3 6
	MHFWD	9 (12) . 7 (5) . (3) 6
	MHMHD	9 (12) . 7 (3) . (1) 6
	FWHMD	9 (11) . 12 . 7 (3) 6
...		

Guide de lecture. La première ligne se lit de la façon suivante : l'union entre l'homme Susuku (Ego) et la femme Mokuri (Alter) participe à 2 circuits matrimoniaux distincts, dont le premier est un mariage avec la mère de l'époux de la mère de l'époux de la fille (DHMHM) ; dans ce circuit, la chaîne de parenté qui relie les conjoints, représentée selon les conventions de la notation positionnelle, descend de l'individu 1 (Ego) à l'individu 6 (la fille d'Ego), passe par l'individu 9 (le mari de la fille d'Ego) pour remonter à l'individu 12 (la mère de 9), passant ensuite à l'individu 7 (le mari de 12) pour remonter à l'individu 3 (la mère de 7).

Tableau 4. Liste synthétique triable

Liste détaillée (triable)				
1	ZD	H()FF	Asanri - Eukiye	451 (452 540) 356 454
2	FDD	H(H)FF	Poto - Apokonini	494 (495) 492 493
3	MSD	H(F)HF	Matato - Numepo	31 (24) 23 29
4	MDD	H(F)FF	Kapai Iputipo - Napanako	315 (316) 314 313
5	FFDD	HH(H)FF	Pesaipo - Piruru	93 91 (90) 63 72
5	FFDD	HH(H)FF	Kamapo - Mori	83 89 (90) 124 87
6	MBD	HF()HF	Kamapo - Atu	83 88 (232 92) 73 80
6	MBD	HF()HF	Kamapo - Mori	83 88 (232 92) 73 87
6	MBD	HF()HF	Inkiman - Marinu	417 419 (547 546) 410 415
7	MMSD	HF(F)HF	Napeta - Paiko	336 314 (316) 332 331
8	FMDSD	HH(F)FHF	Muyopo - Mokoripo	32 39 (214) 24 23 30
9	MMDDD	HF(F)FFF	Pesoro - Iriuna	602 560 (606) 608 559 565
10	FFBDD	HHH()HFF	Kurumuku - Kowaraiye	35 62 64 (296 264) 238 117 118
11	FW	H(H).F	Iwana - Iporika = Ponoriu - Ponoriu	235 (231) . 15
11	FW	H(H).F	Yawikon - Akowani = Yawinapu - Yawinapu	508 (550) . 509
...				

Guide de lecture. La première ligne se lit de la façon suivante : le premier circuit recensé du corpus est le mariage avec la fille de la sœur (ZD en notation standard et H()FF en notation positionnelle) ; les conjoints sont Asanri (Ego) et Eukiye (Alter) ; la chaîne de parenté qui les relie est représentée selon les conventions de la notation positionnelle : elle monte de l'individu 451 (Ego) aux individus 452 et 540 (ancêtres apicaux qui sont eux-mêmes des conjoints) puis descend à l'individu 356 (la sœur d'Ego) puis à l'individu 454 (Alter).

Résultats sous forme de réseaux

Outre ces différents types de listes, Puck produit des réseaux codés lisibles par le logiciel Pajek. En plus des réseaux de parenté représentant un corpus généalogique en tant que tel – où les individus sont représentés par des triangles et des cercles, et les relations de mariage et de filiation par des lignes (Figure 3) – plusieurs types de réseaux sont produits pour l'analyse des structures matrimoniales. Parmi eux, le « réseau matrimonial » se compose de tous les circuits retenus par le recensement matrimonial. L'« armature » de ce réseau est obtenue en réduisant les chaînes de consanguinité à de simples lignes qui apparaîtront en pointillés dans les représentations graphiques^{xvi}. L'armature du réseau matrimonial Trio (en l'occurrence composé des mariages consanguins et des renchéarnements dans les limites du 2^e degré canon) est montrée dans la Figure 4. Les lignes continues représentent les mariages (codés selon les types de circuits dont ils font partie) et les lignes en pointillés représentent les chaînes de consanguinité (les chiffres indiquent le type de circuit matrimonial concerné). Un réseau de ce genre permet de mieux comprendre comment les mariages sont articulés les uns aux autres pour former des « constellations » d'alliances reliées entre elles ; le réseau matrimonial Trio, par exemple, contient une trentaine de telles constellations dont celle qui

rassemble le plus grand nombre d'individus – la constellation maximale - est représentée dans la Figure 4.

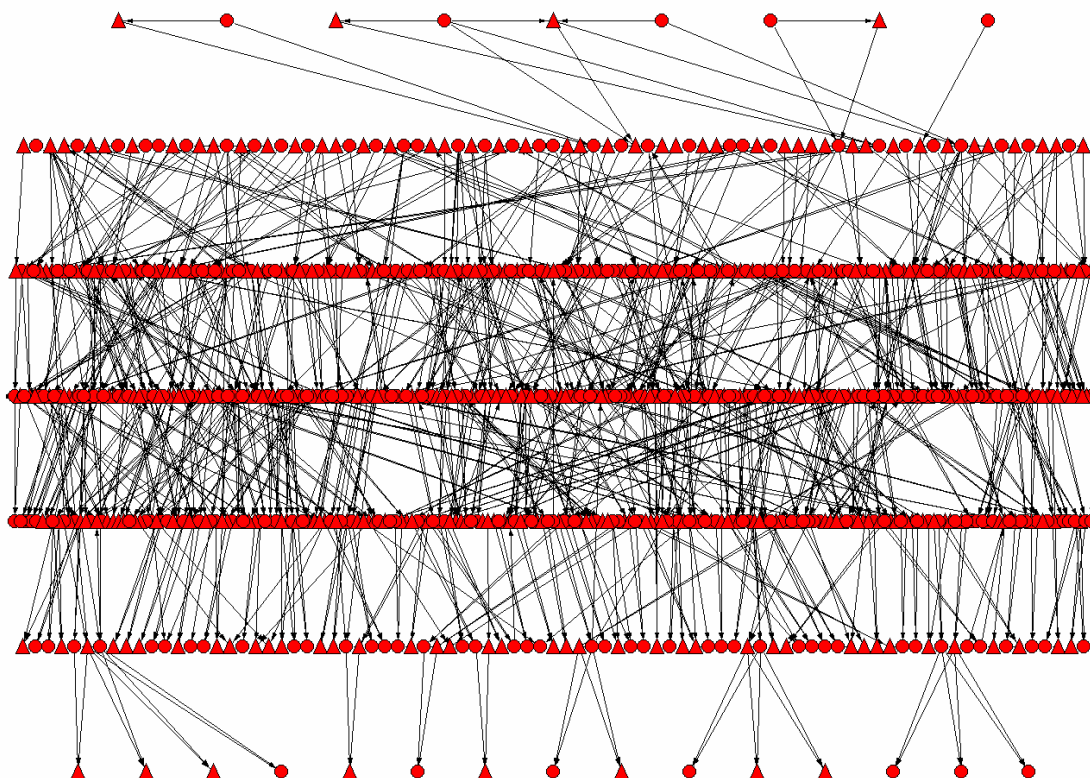


Figure 3. Réseau de 1^e ordre du corpus Trio (affiché sous Pajek)

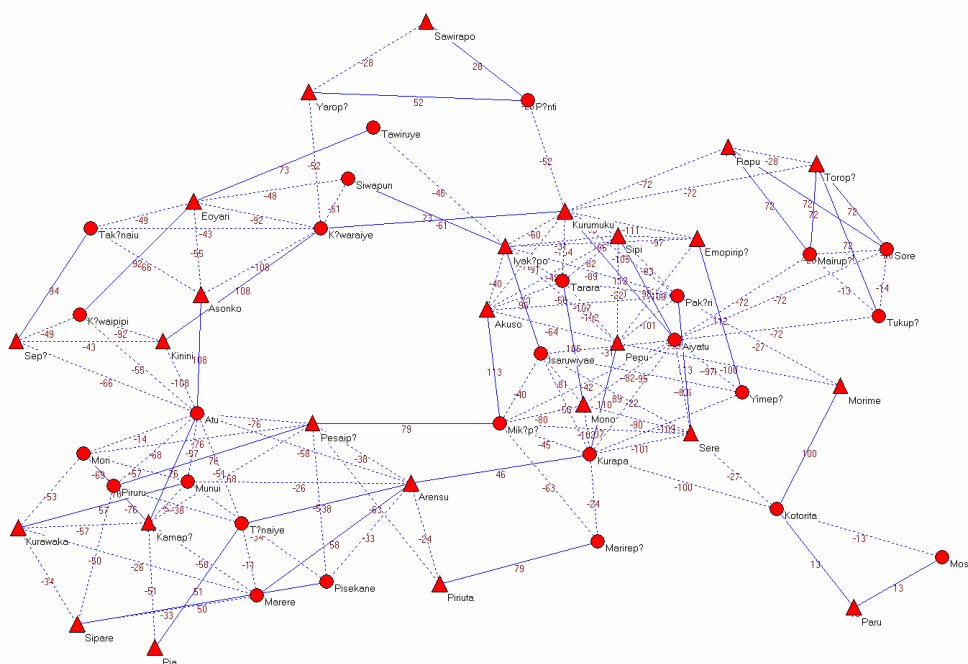


Figure 4. Constellation maximale du réseau matrimonial du corpus Trio (affichée sous Pajek)

Un autre type de réseau analytique que produit Puck est un réseau de « deuxième ordre » (cf. Hamberger *et al.* 2004) dans lequel les points représentent les différents types de circuits – leur taille indiquant les effectifs de ces types – et les valeurs des lignes reliant ces points représentent le nombre de mariages qui font partie des deux types de circuits à la fois (Figure 5). En ce qu'ils permettent de mesurer les effectifs d'unions qui relèvent de deux types de mariage simultanément (unions à la fois avec MBD et avec FFBSD par exemple), les réseaux de deuxième ordre permettent d'évaluer les degrés d'interdépendance entre différents types de mariage recensés au sein du corpus^{xvii}.

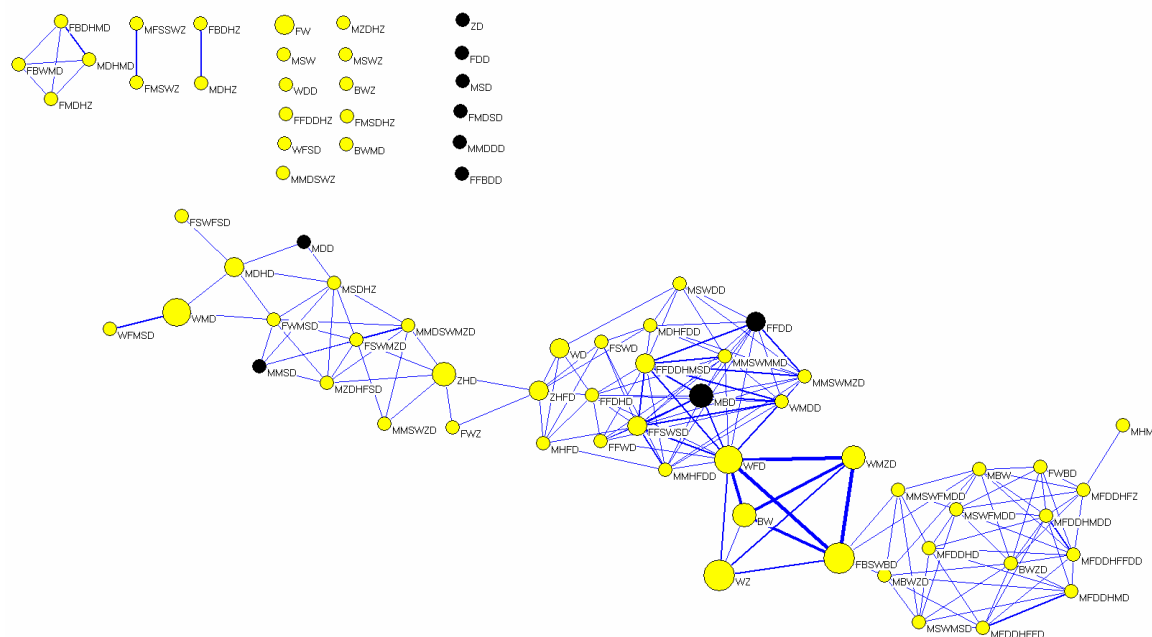


Figure 5. Réseau de 2° ordre affichée sous Pajek représentant les relations entre types de mariages consanguins (cercles foncés) et types de renchaînements de largeur 2 (de hauteur 2) chez les Trio (ici les fréquences des intersections des types sont indiquées par l'épaisseur des lignes).

Recensements partiels

Pour procéder à une analyse affinée de la structure matrimoniale d'un corpus, Puck offre la possibilité de le partitionner, soit selon des critères « endogènes » propres au réseau généalogique et calculés par le logiciel lui-même (nombre de conjoints, taille des groupes de germains, lignées agnatiques ou utérines, etc.), soit selon des paramètres « exogènes » qui figurent dans le corpus (dates d'événements démographiques, lieu de résidence, profession, etc.).

Des partitions de ce genre permettent aussi de restreindre la partie du corpus sur lequel porte la recherche : les individus membres d'un certain lignage, ceux qui sont nés à une certaine époque, ceux dont les quatre grands-parents sont connus, ceux qui ont contracté plusieurs mariages, etc. Cette possibilité de recensements partiels est non seulement essentielle pour des recherches ciblées mais a aussi une importance plus générale. Il peut en effet être méthodologiquement souhaitable de restreindre le corpus afin d'obtenir des résultats significatifs. La qualité d'un corpus dépend non seulement de sa profondeur généalogique moyenne ou de son biais agnatique ou utérin, mais aussi de propriétés structurelles comme sa

densité ou son degré de connectivité. Ainsi, la qualité d'un corpus peut souvent être améliorée si l'on se borne à certaines de ses parties. Les problèmes d'interprétation de chiffres bruts obtenus à partir d'un corpus global biaisé et lacunaire sont moindres si l'on délimite la recherche aux parties du corpus très denses ou très complètes. C'est le cas par exemple lorsque l'on retient la « région » des individus dont on connaît les quatre grands-parents, ou si l'on se limite à la « bicomposante maximale »^{xviii} du réseau. Enfin, il peut être fructueux d'effectuer des recensements sur différentes parties du corpus afin de s'assurer de son homogénéité sur le plan matrimonial. Une telle démarche permet d'éviter de tirer de fausses conclusions alors que le corpus comprend d'importantes lignes de rupture.

Problèmes et techniques d'interprétation des résultats

Comment lire les chiffres ?

Connaître les fréquences des types de mariage au sein d'un corpus généalogique ou d'une de ses parties, et pouvoir repérer la manière dont ils se combinent et se composent, est l'objectif premier de toute analyse de pratiques matrimoniales. Il est toutefois nécessaire de pouvoir évaluer combien ces fréquences sont significatives : dans quelle mesure divergent-elles de celles auxquelles on aurait pu s'attendre si toutes les unions avaient été réalisées de façon aléatoire, indépendamment des mariages antécédents et des liens de parenté induits par ces mariages ?

Soulignons à quel point la nature même de l'objet étudié rend cette tâche complexe. Tout d'abord, il n'y a pas de mariages qui ne dépendent pas, d'une façon ou d'une autre, des autres liens de parenté en présence : le seul fait que les individus se marient plus facilement avec des personnes avec lesquelles il existe déjà un lien direct ou indirect, et que leurs propres horizons matrimoniaux dépendent en partie de ceux de leurs parents (ne serait-ce qu'en raison de leurs lieux de naissance), introduit des inflexions qui tendent à engendrer des circuits au sein du réseau de parenté. Ainsi, l'émergence de bouclages consanguins et de renchéâtements d'alliance au sein du réseau n'a rien de significatif en soi ; leur absence serait plutôt révélatrice de lacunes au sein du corpus. De ce point de vue, c'est avant tout la fréquence *relative* de différents types de circuits les uns par rapport aux autres qui permet d'émettre des jugements sur des évitements et des préférences.

Ensuite, un corpus ne représente jamais qu'une image lacunaire et incomplète du réseau généalogique réel. Ces lacunes ne sont en général pas distribuées de façon aléatoire, mais favorisent certaines relations de parenté plutôt que d'autres. Un tel biais est pratiquement inévitable, même si l'on prend le soin d'obtenir des données aussi équilibrées et complètes que possible, car il arrive que les sources elles-mêmes (orales ou écrites) soient infléchies par le système de parenté en vigueur. Dans une société patrilocale par exemple, il sera plus difficile de repérer les lignes utérines que les lignes agnatiques ; dans une société bilinéaire les chaînes de parenté sexuellement homogènes seront plus facilement mémorisées que les chaînes mixtes ; dans une société cognatique et dravidienne, l'on se rappellera plus facilement des liens horizontaux (passant par les conjoints) que des liens verticaux (passant par les ancêtres), etc. Il est donc nécessaire de pondérer les résultats obtenus en fonction de ces biais incontournables : la présence de beaucoup plus de mariages avec un type de parent « A » qu'avec un type de parent « B » sera à interpréter différemment selon qu'il existe ou non, dans le corpus en question, beaucoup plus de parents de type « A » que de type « B ».

Même dans le cas où les proportions entre les types de parents épousés et les types de parents présents correspondent, l'interprétation des résultats sera différente selon que l'on juge que le biais du corpus est fonction ou non d'une orientation systématique des choix

matrimoniaux. Imaginons un corpus pour lequel un recensement révèle dix fois plus d'unions avec FBD qu'avec MZD, mais où, en même temps, les FBD sont dix fois plus nombreuses que les MZD. Les effectifs relatifs des cousines épousées correspondent donc exactement aux effectifs relatifs des cousines disponibles. Cette différence entre les deux types de mariage (avec la FBD et avec la MZD) est-elle le signe d'une préférence pour les FBD (et d'un évitement des MZD), ou reflète-t-elle tout simplement un biais du corpus ? Cela dépend du nombre de MZD parmi les épouses recensées qui ont échappé au chercheur lors de la phase de collecte des données. Or, la réponse est très différente selon les cas. Un anthropologue qui connaît personnellement les couples de son corpus exclura peut-être qu'il puisse y avoir des liens de consanguinité proche qu'il n'aurait pas repérés : s'il ne connaît pas les grands-mères maternelles des conjoints, dira-t-il, c'est qu'ils n'ont pas une grand-mère maternelle commune, car celle-ci aurait forcément été mémorisée. Autrement formulé, cet argument revient à penser que les comportements matrimoniaux et les biais du corpus sont corrélés, de sorte que le déséquilibre entre les types de parents présents dans le corpus ne fait que confirmer la disproportion entre les types de parents épousés. Mais la situation se présentera tout autrement pour un historien travaillant par exemple à partir de registres paroissiaux. Dans ce cas, un biais agnatique résultant d'une règle de résidence virilocale n'est pas en corrélation avec les régularités matrimoniales. L'absence d'information sur les grands-mères maternelles des conjoints dans le corpus, ne veut en rien dire que les conjoints n'étaient pas des cousins matrilatéraux, voire qu'ils ne se sont pas mariés précisément en raison de ce lien. Dans ce cas, rendre le taux des unions FBD comparable au taux des unions MZD exigerait que l'on multiplie ce dernier par dix : les fréquences obtenues du recensement indiqueraient alors une équiprobabilité de ces deux types de mariage, et le surplus de mariages avec FBD recensés révélerait plutôt des règles de résidence et de filiation qu'une quelconque préférence matrimoniale. Dans la plupart de cas, on se trouve entre ces deux extrêmes. Mais quoi qu'il en soit, l'interprétation des fréquences de mariages en présence d'un déséquilibre important entre les liens paternels, maternels et conjugaux recensés nécessite toujours une argumentation explicite. Seuls peuvent s'en passer les chercheurs qui (comme certains des historiens qui étudient des sources plus récentes, ou des ethnologues de petites communautés cognatiques et endogames) ont la chance de travailler avec un corpus relativement équilibré. Les statistiques produites par Puck permettent de rendre visible le degré de déséquilibre d'un corpus, qui est souvent plus fort qu'on ne le croit.

Enfin, même si le corpus représente une image complète ou du moins homologue au réseau réel, le problème de l'interprétation des chiffres se pose toujours, justement parce qu'il s'agit d'un réseau. Les liens de parenté n'existent pas de façon autonome, mais sont intégrés à des structures réticulaires dans lesquels chaque mariage fait partie d'une constellation incluant plusieurs mariages relevant des générations précédentes. De façon générale, les pratiques matrimoniales des générations précédentes déterminent la distribution des liens de parenté et donc des circuits matrimoniaux qui *peuvent* se former à la génération suivante, avant même que se pose la question de la prise en compte ou non des normes sociales par cette dernière génération. Pour cette raison, la présence d'un circuit matrimonial au sein du corpus, par exemple un mariage du type MBD, ne peut pas être traitée comme un événement indépendant dont la probabilité serait simplement fonction du nombre des relations de parenté correspondantes. Ce nombre n'est pas une donnée exogène, mais dépend lui-même de la configuration des circuits matrimoniaux en présence.

Ces problèmes sont encore loin d'une solution satisfaisante sur le plan théorique comme technique. Le logiciel Puck offre toutefois divers procédés qui apportent certains éléments de réponse.

Evaluation du corpus

L'un d'entre eux consiste en une série de procédures qui produit, sous forme de tableaux et de diagrammes, les principales caractéristiques du corpus : profondeur généalogique, biais agnatique ou utérin, distribution des fratries, etc. Ensemble, ces résultats permettent de dresser un « profil » du corpus. Même s'ils ne sont pas directement utilisables pour « normaliser » des fréquences brutes, ils permettent de relativiser ces fréquences et de fournir le contexte nécessaire à leur interprétation. Ainsi, un corpus avec un important biais agnatique n'est probablement pas apte à démontrer la prévalence du mariage dit « arabe » (avec FBD) ; de même, un corpus avec une très faible profondeur généalogique peut difficilement être utilisé pour révéler l'absence de mariages entre consanguins. Outre cette fonction importante de contrôle, ce « diagnostic » du corpus peut aussi guider la recherche de terrain, en indiquant les lacunes auxquelles il faut remédier et les déséquilibres à compenser.

Les figures 6a-h en donnent un exemple. La Figure 6a montre le nombre d'individus masculins et féminins du corpus Trio dont on connaît soit 1, soit 2, soit 3, jusqu'à 8 arrière grands-parents, et la Figure 6b présente le nombre d'individus masculins et féminins dont on connaît chacun de ces huit arrière grands-parents. La Figure 6c montre le nombre d'individus masculins et féminins ayant un seul enfant, ceux ayant de 2 à 4 enfants, et ceux ayant plus de quatre enfants. La Figure 6d donne une image des biais agnatique et utérin du corpus, en ce qu'elle évalue, pour chaque génération, la proportion des individus ayant un ascendant agnatique ou utérin connu par rapport au nombre d'individus dont on connaît au moins l'un des deux. La Figure 6e présente la fréquence des unions avec les cousins de premier degré, et la Figure 6f montre, pour les mariages consanguins (colonne 2), les unions multiples (colonnes 1 et 4) et les renchaînements par des consanguins de même sexe (colonne 5) ou de sexe différent (colonne 3), la part des unions qui s'accorde avec une logique dravidienne interdisant également le mariage entre générations adjacentes^{xix}.

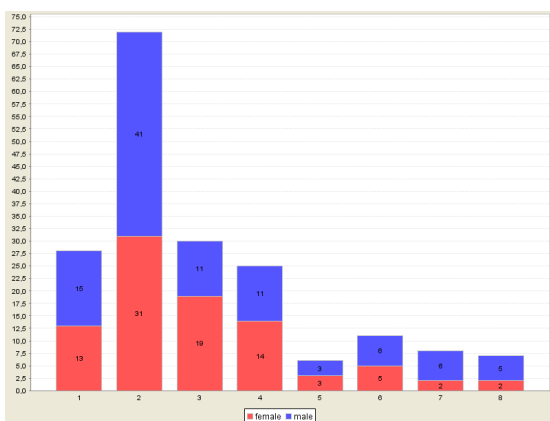


Figure 6a

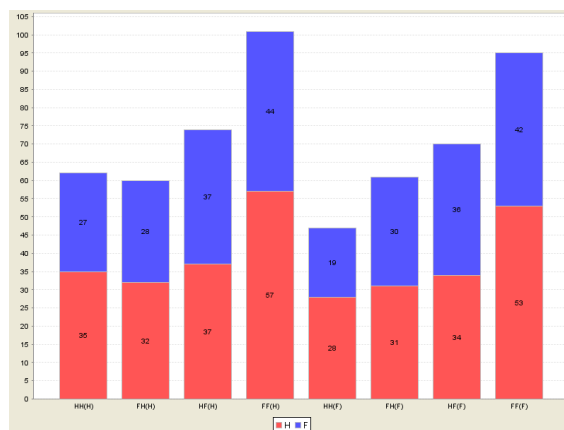


Figure 6b

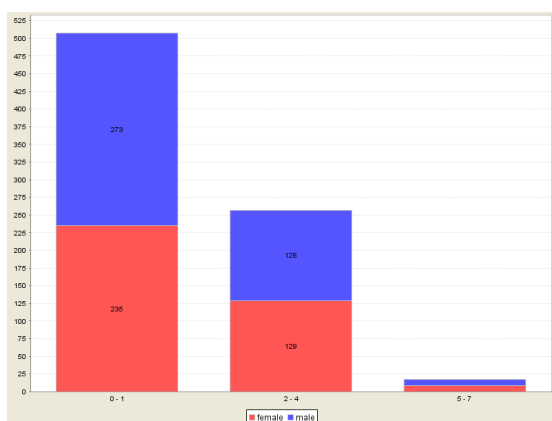


Figure 6c

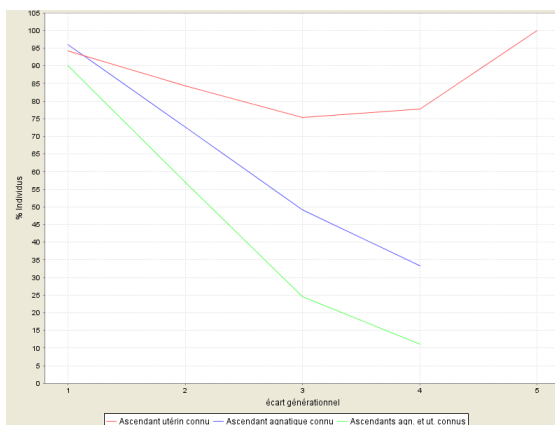


Figure 6d

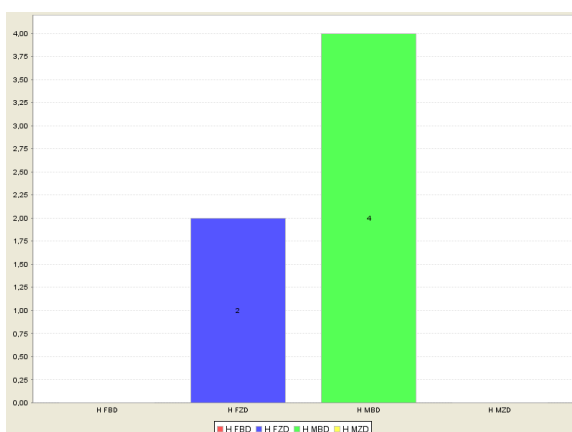


Figure 6e

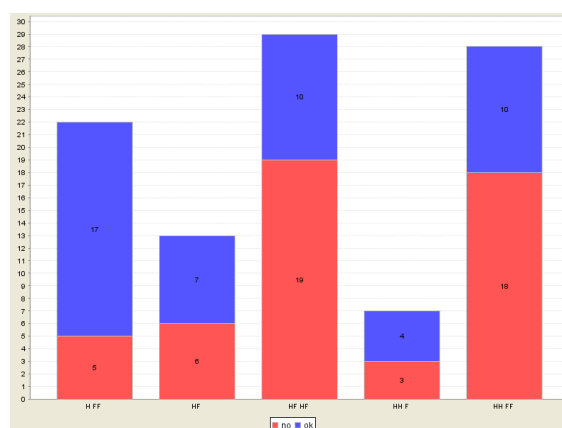


Figure 6f

Techniques de simulation

Un autre type de procédé consiste en l'élaboration de corpus généalogiques simulés pouvant servir de contrôle pour des corpus généalogiques réels. Une approche possible serait de comparer les résultats du recensement avec ceux d'une simulation qui modélise un réseau généalogique idéal, c'est-à-dire sans les biais et les lacunes caractéristiques d'un corpus réel. Il existe des logiciels pour effectuer de telles simulations (par exemple le logiciel GeneoRnd^{xx} de V. Batagelj). Toutefois, cette façon de procéder ne peut être qu'une première approximation. Les réseaux ainsi simulés diffèrent des corpus réels à beaucoup plus d'égards que la seule structure matrimoniale, ce qui réduit de beaucoup la portée des divergences entre les fréquences des circuits matrimoniaux dans les deux réseaux^{xxi}.

Une seconde démarche, plus gratifiante, consiste à engendrer un corpus simulé qui incorpore les mêmes biais que le corpus réel. L'idée est ici de comparer deux réseaux qui ne diffèrent que du point de vue de la structure matrimoniale : le réseau construit partagerait l'ensemble des caractéristiques du réseau original, sauf que les mariages seraient distribués au hasard. Les biais issus des conditions de recherche seraient ainsi neutralisés. Une façon d'obtenir une telle simulation *a posteriori* consiste à permuter tous les mariages du corpus

original (celui que l'on veut tester) de façon aléatoire. Un tel « test de permutation » a été proposé et réalisé par D.R. White à l'aide de son logiciel Pgraph (White 1999). Cette approche pose toutefois certains problèmes. La permutation des liens de mariage induit inévitablement une permutation des liens de filiation, qui, en affectant la taille des fratries et l'homogénéité sexuelle des liens consanguins, modifie considérablement les propriétés du corpus qu'on voulait préserver^{xxii}.

Une troisième démarche consisterait à modéliser, par simulation, non pas un réseau de parenté en soi, mais une enquête de terrain, en choisissant les paramètres de cette enquête (nombre, âge et sexe des informateurs, les probabilités de (re-)mariage, étendu de la mémoire généalogique, etc.) de façon à reproduire les propriétés structurelles du corpus généalogique réel que l'on cherche à étudier. C'est cette nouvelle piste, largement programmatique, qu'introduit le logiciel Puck. Celui-ci incorpore un module d'enquête généalogique simulée qui permet d'engendrer un corpus aléatoire selon des paramètres à sélectionner. La difficulté avec cette option est que le rapport entre les divers paramètres qui entrent en jeu lors d'une enquête (caractéristiques des informateurs, mémoire généalogique, etc.) et les propriétés structurelles du corpus qui en résultent (taille des fratries, nature et le nombre des grands-parents connus, biais agnatique ou utérin, etc.) est très souvent de caractère non-linéaire, et ne se réduit pas à une simple relation biunivoque entre causes et effets singuliers. Le biais agnatique ou utérin d'un corpus, par exemple, est souvent le résultat conjoint de la mémoire généalogique des individus, de la structure résidentielle, de l'étendue de l'aire matrimoniale et du sexe des informateurs.

Si l'association d'un module de contrôle et d'un module de simulation ne peut constituer une solution définitive aux problèmes que soulève l'interprétation des fréquences des circuits matrimoniaux, elle permet toutefois de mieux évaluer la validité des modèles et d'assurer la comparabilité des corpus.

Conclusion

L'objectif du logiciel Puck est de permettre une analyse systématique des corpus généalogiques, afin de les mettre en perspective avec les normes matrimoniales, les terminologies de parenté, et d'autres institutions caractérisant le système de parenté d'une société. Sans incarner une orientation interprétative particulière, il se veut un outil polyvalent pouvant être mis au service d'une variété d'approches différentes. C'est ainsi qu'il peut contribuer à relancer les débats théoriques inéluctablement abandonnés face à l'absence de moyens permettant le traitement systématique des pratiques de mariage.

Certes, la théorie de la parenté est beaucoup plus qu'une théorie des pratiques matrimoniales, tout comme la parenté elle-même est beaucoup plus qu'un système symbolique pour représenter et organiser des structures généalogiques. C'est également un langage dans lequel s'expriment des structures sociales, des systèmes cosmologiques, des rapports de pouvoir, des équations logiques. Mais ce sont les relations généalogiques et matrimoniales qui fournissent à ce langage ses moyens d'expression, et qui constituent son contenu paradigmatique. Une recherche sur les phénomènes de parenté qui ne passe pas par une étude des réseaux généalogiques et des pratiques matrimoniales risque de perdre, sinon la base empirique des modèles qu'elle construit ou reconstruit, tout du moins un moyen important pour saisir leur validité et leur portée.

Puck constitue avant tout un outil de critique et d'évaluation. Son usage peut rendre le chercheur plus indépendant des points de vue nécessairement partiels et souvent partiels de ses informateurs, mais aussi de ses propres préjugés théoriques qui pèsent d'autant plus si l'analyse du corpus se limite au comptage d'un petit nombre de types de mariage supposés significatifs. En même temps, il facilite le repérage des lacunes et des biais du corpus,

permettant au chercheur soit d'améliorer la base empirique de ses jugements, soit de relativiser les conclusions qui peuvent être tirés à partir de cette base.

Toutefois, l'utilité d'outils informatiques dans les études de parenté dépend de deux conditions. La première concerne le libre accès aux corpus généalogiques employés afin que la qualité des données puisse être évaluée et les analyses reproduites par d'autres^{xxiii}. La seconde se rapporte à un savoir-faire, suffisamment partagé pour identifier les conclusions erronées et les hypothèses infondées que peuvent receler les analyses informatiques. Plus la base empirique des analyses restera soustraite à l'évaluation publique et plus le cercle de chercheurs ayant les compétences pour les critiquer sera restreint, plus alors les logiciels risqueront de devenir des instruments idéologiques. C'est seulement dans un environnement qui ne considère pas les données généalogiques comme un capital symbolique à conserver jalousement, et où la formalisation n'est pas ressentie comme un obstacle, que les outils informatiques comme Puck peuvent contribuer à une meilleure compréhension des pratiques qui engendrent les réseaux de parenté.

Bibliographie

- Barry, L. (1996), *La parenté recomposée. Figures peuls de l'alliance sur les hauts plateaux de l'Adamaoua (Nord Cameroun)*, Thèse de doctorat, Université Paris X- Nanterre.
- Barry, L. (1998), « Les modes de composition de l'alliance. Le 'mariage arabe' », *L'Homme* 147 : 17-50.
- Barry, L. (2004), « Historique et Spécificités Techniques du Programme Genos », *Ecole « Collecte et traitement des données de parenté »*, <http://llacan.vjf.cnrs.fr/SousSites/EcoleDonnees/extras/Genos.pdf>.
- Batagelj, V. et Mrvar, A. (1998), « Pajek -- A Program for Large Network Analysis », *Connections*, 21(2), 47-57.
- Batagelj, V. et Mrvar, A. (2008), « Analysis of Kinship Relations with Pajek », *Social Science Computer Review*, 26 (2), 224-246
- Brudner, L. et White, D. W. (1997), « Class, Property and Structural Endogamy: Visualizing Networked Histories », *Theory and Society* 26 : 161-208.
- Cazès, M.-H. et Guignard, E. (1991), « Les cercles d'alliance chez les Dogon islamisés de Tabi », in F. Héritier-Augé et E. Copet-Rougier (éds), *Les complexités de l'alliance, II : Les systèmes complexes d'alliance matrimoniale*. Paris, Éditions des Archives contemporaines
- Collard, C. (1994), « Preference and Limit of the Preference : The case of Marriages Between Two Brothers and Two Sisters in French Quebec », Communication presented at the EASA conference, Oslo.
- Coult, A.D. et Randolph, R.R. (1965), « Computer Methods for Analyzing Genealogical Space », *American Anthropologist*. 67 : 21-9.
- Delille, G. (1985), *Famille et propriété dans le Royaume de Naples (XVe-XIXe siècles)*. Rome-Paris, Ecole Française de Rome.
- De Nooy, W., Mrvar, A., et Batagelj, V. (2005), *Exploratory Social Network Analysis with Pajek*, New York, Cambridge University Press.
- Ferchiou, S. (2000), « Marriage "arabe" et alliance redoublée dans la société tunisoise », in J.-L. Jamard, E. Terray et M. Xanthakou (éds), *En substances. Textes pour Françoise Héritier*. Paris, Fayard, pp. 183-194.
- Gasperoni, M. (2008), *Popolazione, famiglie, parentela a San Marino in età moderna*, Quaderni del Centro Sammarinese di Studi Storici, Università di San Marino

- Grange, C. (2005), "Les réseaux matrimoniaux de l'upper class israélite parisienne à la fin du XIXe siècle", *Annales de Démographie Historique*, 2005/1, pp.131-156.
- Hamberger, K., Houseman, M., Daillant, I., White, D. et Barry, L. (2004), « Matrimonial Ring Structures », *Mathématiques, Informatiques et Sciences Humaines. Journal de l'Ecole des Hautes Études en Sciences Sociales* 168, Hiver 2004, Spécial *Les Réseaux Sociaux*, pp. 83-119.
- Héritier, F. (1974), « Systèmes Omaha de parenté et alliance. Etude sur ordinateur du fonctionnement réel d'une société africaine », in P.A. Ballonoff (éd.) *Genealogical Mathematics*, Paris-La Haye, Mouton.
- Houseman, M. et White, D.R. (1996), « Structures réticulaires de la pratique matrimoniale », *L'Homme* 139 : 59-85
- Houseman, M. et White, D.R. (1998), « Network Mediation of Exchange Structures: Ambilateral Sidedness and property Flows in Pul Eliya (Sri Lanka) », in T. Schweizer et D.R. White (éds.) *Kinship, Networks and Exchange*. Cambridge, Cambridge University Press.
- Jolas, T., Verdier, Y. et Zonabend, F. (1970), « Parler famille », *L'Homme* X (3), 5-26.
- Richard, P. (1993), « Étude des renchaînements d'alliance », *Mathématiques, Informatique et Sciences humaines* 123 : 5-35.
- Segalen, M. (1985), *Quinze générations de Bas-Bretons. Parenté et société dans le pays bigouden sud. 1720-1980*. Paris, PUF.
- Selz, M. (1994), « Traitement informatique de données généalogiques : le logiciel GENPAR », *L'Homme* 130 : 129-136.
- Sinibaldi, S. (2007): « Réseaux des liens matrimoniaux. Le choix du conjoint dans une commune toscane au XX^e siècle », *L'Homme* 181 : 41-74
- White, D.R. et Jorion, P. (1992), « Representing and Computing Kinship: A New Approach », *Current Anthropology*, Vol. 33, No. 4. (Aug. - Oct., 1992), pp. 454-463.
- White, D.R., Batagelj, V. et Mrvar, A. (1992), « Analyzing Marge Kinship and Marriage Networks with Pgraph and Pajek », *Social Science Computer Review* 17(3): 245-274.
- White, D.R. (1997), « Structural Endogamy and the Graphe de Parenté », *Mathématiques, Informatique et Sciences Humaines* 137 : 107-125.
- White D.R. (1999), « Controlled Simulation of Marriage Systems », *Journal of Artificial Societies and Social Simulation* 2 (3), <http://jasss.soc.surrey.ac.uk/2/3/5.html>
- White, D. R. et Johansen, U. (2004), *Network Analysis and Ethnographic Problems: Process Models of a Turkish Nomad Clan*. Boston, Lexington Press.

ⁱ Le logiciel Puck (*Program for the Use and Computation of Kinship data*), écrit en Java 1.6, a été développé en 2007 par Klaus Hamberger au sein du groupe TIP (« Traitement Informatique de la Parenté ») dans le cadre du projet « Traitement informatique des phénomènes de parenté en anthropologie et en histoire : une approche intégrée » (TIPP) dirigé par Michael Houseman et Cyril Grange, et financé par l'Agence Nationale de la Recherche (NT05-4_43503). Par ailleurs, le développement de ce logiciel a bénéficié de discussions avec Laurent Barry, Vladimir Batagelj et Andrej Mrvar. Puck peut être librement téléchargé sur le site www.kintip.net.

Ce texte a été écrit par les trois signataires. Toutefois, la réflexion qui a abouti au logiciel Puck et donc à cet article a été menée collectivement avec d'autres membres actifs du groupe TIP : Elisabeth Anstett, Pascale Constans, Isabelle Daillant, Laurent Gabail, Michael Gasperoni, Sophie Goedefroit et Olivier Kyburz. Les auteurs tiennent à remercier les autres membres du projet, ainsi que Laurent Barry, Dominique Casajus, Eric Jolly, Marianne Lemaire et les lecteurs anonymes de la revue *L'Homme* pour leurs commentaires utiles.

ⁱⁱ Un recensement de tous les mariages entre consanguins et affins dans les limites du 3^e degré canon dure entre 0.5 secondes pour des corpus de moyenne taille (env. 5000 individus) et de faible densité (moins de 10 consanguins par individu marié) et 20 secondes pour des corpus très grands (env. 30000 individus) ou très denses (plus de 200 consanguins par individu marié). Les résultats peuvent être étudiés au moyen de divers outils analytiques et visualisés sous la forme de diagrammes ou de réseaux. Puck propose un formulaire de saisie des données mais permet aussi d'importer des corpus déjà établis dans d'autres formats (tableaux, bases de données, réseaux, GEDCOM), de les analyser, de les manipuler et de les segmenter, puis de les réexporter dans tous ces formats. Un guide d'emploi détaillé est disponible sur www.kintip.net.

ⁱⁱⁱ Nous adoptons ici la notation devenue conventionnelle dans laquelle les chaînes de parenté sont désignées au moyen d'une série de lettres renvoyant aux termes qui sont utilisés en anglais pour signifier les liens de parenté qui composent ces chaînes : F (*father*), M (*mother*), B (*brother*), Z (*sister*), H (*husband*), W (*wife*), S (*son*), D (*daughter*). Rappelons que l'intérêt de cette notation est qu'elle permet de désigner une relation entre Ego et Alter en partant non pas d'Alter comme c'est le cas en français (« la fille du frère de la mère »), mais d'Ego (« Mother's Brother's Daughter »).

^{iv} La « taille » d'un circuit matrimonial comprend plusieurs dimensions, telle que « hauteur » (longueur de la plus longue chaîne linéaire, autrement dit le degré canon maximal des chaînes consanguines) et « largeur » (nombre de mariages qu'intègre le circuit).

^v Indépendamment du développement de Puck, une version préliminaire de Genos sous Java a été développée à la demande du projet TIPP par Murat Ahat.

^{vi} Pgraph, Par-calc et d'autres logiciels de Douglas White sont disponibles sur le site <http://eclectic.ss.uci.edu/~drwhite/pgraph/p-graphs.html>.

^{vii} Le logiciel Pajek (disponible à <http://pajek.imfm.si/doku.php>), développé par Vladimir Batagelj et Andrej Mrvar, est utilisé aujourd'hui par un grand nombre de chercheurs en sciences sociales (voir Batagelj et Mrvar 1998 ; de Nooy, Mrvar et Batagelj 2005).

^{viii} Une quarantaine de macro-instructions pour Pajek (*Tip4Pajek*), développée par Klaus Hamberger, est disponible sur les sites <http://kintip.net> et http://intersci.ss.uci.edu/wiki/index.php/TIPP_Kinship_and_computing.

^{ix} Un ascendant « apical » est un ascendant qui n'a pas lui-même d'ascendant.

^x Pour simplifier, la Figure 2a représente le lien de pleine germanité par une ligne directe. Toutefois, le logiciel ne traite pas la germanité comme une relation élémentaire, mais comme une relation complexe impliquant deux chaînes qui passent, respectivement, par le père et par la mère.

^{xi} Puck offre la possibilité de prendre en compte ou non de tels circuits englobants.

^{xii} Le degré « canon » correspond au nombre maximal de générations reliant deux personnes à un ancêtre commun ; par exemple, la relation entre Ego et la fille du frère du père de son père (FFBD ou HHH()HF) est de 3^e degré canon. Par contre, le degré « civil » correspond au nombre de liens (de consanguinité et d'affinité) reliant les personnes ; par exemple, la relation entre Ego et la fille du frère du père de son père est de 5^e degré civil.

^{xiii} La « caleulette » de Puck permet d'afficher pour une relation de parenté sélectionnée : ses représentations alternatives, sa traduction en plusieurs notations, son degré (civil et canon), le nombre de liens d'affinité et de composantes consanguines qu'elle intègre, ainsi que le profil de ces composantes selon l'écart générationnel et le degré d'obliquité qu'elles impliquent, selon la nature de leur lien de germanité apical, selon que les individus terminaux des composantes sont de même sexe ou de sexe différent, selon le caractère agnatique, utérin, cognatique ou bilatéral (pleine germanité) du lien qui unit ces individus, selon le nombre de

changements de sexe que comporte ce lien et selon son caractère parallèle ou croisé dans une logique dravidienne.

^{xiv} Le corpus Trio est disponible en format lisible par Puck et Pajek sur <http://kinsource.net/kinsrc/bin/view/KinSources/Trio>.

^{xv} Toutes les propriétés des circuits mentionnés dans la note 12 peuvent être utilisées pour caractériser les circuits listés dans le rapport de recensement. Il est également possible d'afficher le pourcentage que représente chaque type de circuit (par exemple, unions avec MBD) par rapport à l'ensemble des relations de parenté correspondantes dans le corpus (l'ensemble des relations MBD), et de calculer sur ces bases une première approximation d'un « coefficient de préférence ».

^{xvi} Le concept d'« armature » d'un réseau de parenté (donc d'un réseau où les liens de consanguinité sont représentés par des simples lignes plutôt que par les chaînes) s'applique de la même façon à un circuit singulier et à un réseau matrimonial entier (composé d'un ensemble de circuits).

^{xvii} Puck permet également une décomposition progressive automatisée de ces réseaux de 2ème ordre (voir Hamberger et al. 2004) afin de repérer le poids relatif de différents types de circuits dans l'émergence d'un profil matrimonial particulier.

^{xviii} Une bicomposante matrimoniale d'un réseau de parenté correspond grossièrement à un ensemble d'individus dont les mariages font partie de circuits matrimoniaux. Ces sous-ensembles et les « ponts » généalogiques qui les relient forment le « noyau » (Houseman et White 1996) du réseau et représentent sa partie fortement interconnectée, répondant au critère d'endogamie structurelle (White 1997). La bicomposante matrimoniale maximale est la bicomposante matrimoniale la plus importante en termes d'effectif.

^{xix} On discernera par ailleurs certaines tendances communes aux systèmes dravidiens permettant le mariage avec la nièce utérine : le biais utérin du réseau, le poids relatif des ascendants utérins connus, la dominance de mariages avec la cousine croisée matrilatérale, et la forte présence de mariages intergénérationnels.

^{xx} Disponible à <http://vlado.fmf.uni-lj.si/pub/networks/pajek/howto/geneoRnd.htm>

^{xxi} L'idée de simuler un réseau de parenté complètement aléatoire ayant valeur de contrôle – où chaque individu de sexe différent aurait la même chance d'épouser chacun des autres – est irréaliste sinon absurde. Ne serait-ce que pour tenir compte des écarts générationnels et spatiaux par exemple, il est nécessaire d'introduire une certaine conception d'horizon matrimonial dans la construction même d'un tel réseau « aléatoire ».

^{xxii} Dans un corpus réel, la taille des fratries présente une distribution très inégale (pour certains couples on connaît tous les enfants, pour d'autres on ne connaît aucun enfant), et les lignées de descendance sexuellement homogènes (père du père du père, mère de la mère de la mère etc.) sont souvent plus fréquents que des lignes mixtes (père de la mère du père etc.).

^{xxiii} Par exemple, l'archive en ligne établi par Michael Fischer et autres (kinsource.net).