

<b>Acronyme</b>	<b>SimPa</b>		
<b>Titre du projet en français</b>	<b>Simulations de parenté : modélisation de la dynamique matrimoniale et mnémonique dans les réseaux de parenté</b>		
<b>Titre du projet en anglais</b>	<b>Kinship simulations : modelling the dynamics of marriage and memory in kinship networks</b>		
<b>Axe(s) thématique(s)</b>	<input type="checkbox"/> Sciences pour l'ingénieur <input type="checkbox"/> Sciences de la Terre et de l'Univers <input type="checkbox"/> Biologie – Santé <input type="checkbox"/> Agronomie – Ecologie – Environnement <input checked="" type="checkbox"/> <b><u>Dynamiques Humaines et Sociales</u></b> <input type="checkbox"/> Autres		
<b>Méthodes</b>	<input checked="" type="checkbox"/> <b><u>Méthodes mathématiques</u></b> <input type="checkbox"/> Physique statistique <input type="checkbox"/> Informatique fondamentale <input type="checkbox"/> Modélisation et Masses de données		
<b>Type de recherche</b>	<input checked="" type="checkbox"/> <b><u>Recherche Fondamentale</u></b> <input type="checkbox"/> Recherche Industrielle <input type="checkbox"/> Développement Expérimental		
<b>Aide totale demandée</b>	349 046 €	<b>Durée du projet</b>	36 mois

<b>1. CONTEXTE ET POSITIONNEMENT DU PROJET .....</b>	<b>2</b>
1.1. Contexte et enjeux économiques et sociétaux.....	3
1.2. Positionnement du projet .....	4
<b>2. DESCRIPTION SCIENTIFIQUE ET TECHNIQUE .....</b>	<b>5</b>
2.1. État de l'art .....	5
2.2. Objectifs et caractère ambitieux/novateur du projet .....	8
<b>3. PROGRAMME SCIENTIFIQUE ET TECHNIQUE, ORGANISATION DU PROJET.....</b>	<b>10</b>
3.1. Programme scientifique et structuration du projet .....	10
3.2. Management du projet.....	12
3.3. Description des travaux par tâche .....	13
3.3.1 Tâche 1 – Collecte et codage des données [Responsable : CEMAF, INED]	13
3.3.2 Tâche 2 – Modélisation [Responsable : INED]	14
3.3.3 Tâche 3 – Simulation et validation [Responsable : CAMS]	18
3.3.4 Tâche 4 – Implémentation, Valorisation, usages & services	21
3.4. Calendrier des tâches, livrables et jalons.....	21
<b>4. STRATEGIE DE VALORISATION DES RESULTATS ET MODE DE PROTECTION ET D'EXPLOITATION DES RESULTATS .....</b>	<b>22</b>
<b>5. ORGANISATION DU PARTENARIAT .....</b>	<b>23</b>
5.1. Description, adéquation et complémentarité des partenaires.....	23
5.1.1 Partenaire 1 : Institut National d'Etudes Démographiques (chef d'équipe : François Héran)	23
5.1.2 Partenaire 2 : Centre d'Analyse et de Mathématique Sociale (Chef d'équipe : Camille Roth)	24

5.1.3	Partenaire 3 : Centre d'Etudes des Mondes Africaines (Chef d'Equipe : Michael Houseman)	25
5.2.	Qualification du coordinateur du projet .....	27
<b>6.</b>	<b>JUSTIFICATION SCIENTIFIQUE DES MOYENS DEMANDES .....</b>	<b>27</b>
6.1.	Partenaire 1 : INED .....	27
6.2.	Partenaire 2 : CAMS .....	28
6.3.	Partenaire 3 : CEMAF .....	29
<b>7.</b>	<b>ANNEXES.....</b>	<b>29</b>
7.1.	Références bibliographiques .....	29

## **1. CONTEXTE ET POSITIONNEMENT DU PROJET**

Les réseaux de parenté évoluent comme des systèmes dynamiques. La position relative de deux personnes dans le réseau, c'est-à-dire les chaînes formées par des mariages antérieurs et/ou par des liens de descendance qui en découlent, influe sur la probabilité d'un mariage entre ces deux personnes. Cette incidence peut être explicite (prescriptions ou prohibitions de mariage entre certains types de parents ou d'affins), implicite (préférences et stratégies), directe (en faisant intervenir les liens de parenté entre des conjoints potentiels) ou indirecte (par l'intermédiaire de facteurs résidentiels ou autres, qui sont à leur tour corrélés avec certaines relations de parenté). Dans tous ces cas, certaines configurations cycliques – « circuits matrimoniaux » – apparaissent plus fréquemment que d'autres dans le réseau de parenté.

La distribution des fréquences des différents types de circuits matrimoniaux se présente ainsi comme une clé à la compréhension des comportements matrimoniaux sous-jacents. Les études de parenté – en anthropologie, sociologie ou histoire – ont en effet toujours procédé en comptant les occurrences de certains types de circuits. Jadis restreinte à un petit nombre de circuits « typiques » (les mariages entre cousins par exemple), cette recherche peut aujourd'hui, grâce à des nouveaux outils informatiques, être étendue en peu de temps à des centaines de types de configurations dans des réseaux comptant plusieurs dizaines de milliers d'individus. En même temps, de plus en plus de corpus généalogiques de grande taille deviennent accessibles sous forme informatisée. Toutefois, cette nouvelle richesse de données fait apparaître un problème profond qui a longtemps été naïvement négligé : la *fréquence* relative d'un certain type de mariage ne suffit pas pour fournir l'indice d'une *préférence* relative. Or il n'existe jusqu'à présent aucune méthode permettant d'inférer, à partir de la distribution des circuits, les comportements matrimoniaux qui l'ont générée. Les études de parenté se trouvent donc aujourd'hui dans la situation paradoxale où le progrès des outils informatiques les a privées d'un instrument défectueux – le simple comptage de circuits matrimoniaux – sans les munir d'un instrument plus performant. Face à l'immense richesse des données nouvellement accessibles, elles n'ont pas les moyens de les interpréter. Le premier objectif du projet SimPa consiste à débloquent cette situation.

Les circuits matrimoniaux ne représentent pas des événements indépendants dont la probabilité pourrait se calculer à partir de la fréquence des chaînes de parenté impliquées. Cette fréquence dépend elle-même des circuits dans le réseau, de sorte que chaque mariage

d'un type donné change la probabilité de chaque type de mariage. Qui plus est, l'immense majorité de circuits matrimoniaux se forment de façon mécanique par la composition d'autres circuits, et représentent donc des artefacts du réseau plutôt qu'un indice comportemental en soi. Pour pouvoir comprendre les règles comportementales qui engendrent un certain profil de circuits matrimoniaux, il ne suffit pas de considérer chaque type de circuit isolément. Il faut reproduire la morphologie *globale* du réseau, ce qui implique forcément des techniques de simulation. Cette modélisation de l'émergence de différents types de circuits dans un réseau de parenté exige une compréhension approfondie de la formation de cycles dans des graphes faiblement acycliques.

Tout en visant un verrou d'importance éminemment pratique dans les études de parenté, le projet s'attaque donc à une problématique plus fondamentale de la théorie des graphes liée à la formation de la structure des réseaux sous la donnée de certains comportements individuels et locaux (soit la « morphogenèse »), dont le traitement promet des répercussions sur d'autres domaines de l'analyse des réseaux sociaux, voire des réseaux en général.

Le problème de l'interprétation des motifs dans les réseaux de parenté manifeste un aspect supplémentaire : les réseaux auxquels ont affaire les anthropologues et historiens ne sont que des représentations partielles et lacunaires du réseau de parenté réel. La distribution des circuits dans ces réseaux traduit en partie des contraintes mnémoriques (y compris les biais de recherche). La simulation des réseaux de parenté exige donc la modélisation de la mémoire généalogique (*genealogical recall*) aussi bien que celle des comportements matrimoniaux. Cette tâche exige le développement des techniques de reconstruction des réseaux sous l'hypothèse de données manquantes et la mise en place d'outils statistiques nouveaux concernant l'estimation de la présence de motifs dans les graphes et, surtout, de leur représentativité.

Si le projet s'intéresse donc particulièrement à certains motifs propres aux réseaux de parenté, il permettra plus largement, du point de vue théorique, de s'attaquer en toute généralité à des problèmes mathématiques et statistiques pour lesquels la littérature actuelle reste encore très limitée. Les spécificités structurelles des réseaux de parenté les qualifient comme champ expérimental pour le développement des nouvelles méthodologies d'importance pour l'analyse des réseaux en général, et des réseaux sociaux en particulier.

### **1.1. CONTEXTE ET ENJEUX ECONOMIQUES ET SOCIETAUX**

Le projet a pour but pratique de fournir aux chercheurs d'une large gamme de disciplines travaillant avec des corpus généalogiques (anthropologie, histoire, démographie, sociologie, études des migrations, génétique, etc.) des instruments fiables et pertinents pour interpréter et évaluer leurs données. Les apports sociaux et économiques des résultats du projet sont donc surtout indirects – ce sont les enjeux des recherches que les instruments produits rendront possibles. Ceci vaut encore plus des avancées méthodologiques attendues dans le domaine de la théorie des graphes, de la statistique des événements réticulaires et de la simulation sous contrainte de données manquantes. A cet égard, le champ d'applications

possibles coïncide en fin de compte avec celui de l'analyse des réseaux sociaux. Tout en relevant de la recherche fondamentale, le projet se fixe comme objectif de résoudre un problème concret qui verrouille jusqu'à présent tout un champ de recherche concernant les réseaux empiriques. C'est en situant ce problème dans un contexte mathématique plus large que cette recherche envisage d'arriver à une solution fertile pour le traitement d'autres problèmes du même type (relation entre comportement et structure globale, sous contraintes fortes, reconstruction de données de réseaux empiriques et complétion de corpus, notamment).

## 1.2. POSITIONNEMENT DU PROJET

Le projet SimPa se situe dans la perspective du développement de méthodes mathématiques pour modéliser la dynamique des réseaux sociaux.

La modélisation des réseaux en général et des réseaux sociaux en particulier occupe actuellement de nombreuses équipes, au sein de projets financés au niveau européen et international qui, à partir de différents terrains – communautés en ligne, réseaux de collaboration, réseaux pair-à-pair et IP, réseaux urbains, etc. – visent essentiellement à rendre compte de motifs structurels (communautés, particularités topologiques) et à modéliser leur formation en tant que systèmes complexes.

SimPa s'inscrit dans la lignée de ces efforts, en s'en distinguant par trois aspects essentiels et jusqu'ici largement inexplorés :

(1) L'exhaustivité du corpus empirique n'est généralement pas remise en cause, et le biais potentiellement induit n'est généralement pas étudié, en dehors des études de cas très ponctuelles mentionnés dans l'état de l'art (section 2.1). SimPa se situe en amont des projets s'appuyant sur des graphes de terrain en ambitionnant de résoudre de manière générique la question de la reconstruction des données réelles partielles et/ou biaisées. SimPa se rapproche par ailleurs de projets de métrologie de graphes développés en informatique, notamment dans le cadre des réseaux IP (même si ces derniers travaux s'appuient sur des protocoles de collecte relativement particuliers comme l'envoi de sondes sur les réseaux).

(2) Bien que de très gros efforts ont été menés autour de la détection de motifs (CFinder, Toscana, MFinder, etc.), la question de la corrélation entre fréquences d'apparition de motifs reste souvent inexplorée. La modélisation de SimPa est centrée sur la question de l'apparition conjointe de certains motifs et l'induction potentielle d'artefacts d'observation, et non pas sur celle de la corrélation entre certains effets contribuant à la dynamique de création de liens. C'est le cas notamment des équipes travaillant autour de la plate-forme SIENA, par ailleurs centrées sur l'étude de réseaux complets et ne comportant donc pas d'enjeux liés à la reconstruction. SimPa ne vise effectivement pas à définir ou détecter des motifs, mais plutôt à détecter en quoi des motifs connus sont ou non significatifs, tout en prenant en compte les corrélations dues aux effets de réseau.

(3) Le cadre des réseaux de parenté, très contraints, constitue un cadre original et privilégié pour résoudre ces problèmes. Souvent abordés comme des réseaux ego-centrés (par exemple le projet KASS sur le rôle des réseaux familiaux dans l'assistance mutuelle), leurs propriétés en tant que réseaux d'interaction multicentriques, faisant intervenir l'interdépendance d'une pluralité d'événements, ont fait l'objet de peu de travaux. Les rares projets s'interrogeant sur les propriétés émergentes des réseaux de parenté, celui de D. R. White et F. Harary sur les phénomènes de cohésion (NSF 2002), celui de l'équipe TIPP (ANR 2005), ont souligné le besoin de travaux sur le fonctionnement de ces réseaux en tant que systèmes complexes, fondés sur une modélisation mathématique et statistique adéquate. C'est une telle modélisation que propose SimPa.

## **2. DESCRIPTION SCIENTIFIQUE ET TECHNIQUE**

### **2.1. ÉTAT DE L'ART**

Les études de parenté ne se sont que récemment doté des moyens informatiques nécessaires pour une analyse empirique des pratiques matrimoniales, pourtant fondement de tout système de parenté. Bien que l'ordinateur ait été introduit depuis les années 1970 dans ce champ de recherche (Coult et Randolph 1965 ; Gilbert et Hammel 1966 ; Héritier 1974), et qu'un nombre croissant de travaux témoigne du recours aux outils informatiques (voir notamment : Héritier 1981, Delille 1985, Segalen et Richard 1986, Cazes et Guignard 1991, Selz 1994, Houseman et White 1996, Barry 1998, White et Johanson 2004), la plupart des logiciels disponibles ne permettaient que de rechercher certains types de mariages spécifiés au préalable. Cette situation a changé avec le développement de logiciels permettant de recenser la totalité des circuits matrimoniaux dans un réseau de parenté : *Par-Calc* de Douglas White (1992-1996, voir White et Jorion 1992), *Genos* de Laurent Barry (1996-1997, voir Barry 1996, 2004), et plus récemment *Puck* de Klaus Hamberger (2007-2009, voir Hamberger, Houseman et Grange 2009). Nous disposons désormais des moyens de repérer et de compter des milliers de circuits matrimoniaux de centaines de types différents.

Paradoxalement, ce nouvel accès à des données quantitatives infiniment plus riches et complètes qu'à l'époque classique des décomptages « à la main » pose les études de parenté devant des problèmes méthodologiques jusqu'alors naïvement négligés. Tant que les analyses se bornaient à un petit nombre de types de configurations supposés importants pour le comportement matrimonial, types dont la distribution était supposée aléatoire, les fréquences relatives des configurations de mariage ont pu être appréhendées comme fournissant une image simple et directe des règles matrimoniales d'une société. Dans maints travaux, des décomptages de mariages entre certains types de parents (notamment entre les quatre types de cousins germains) sont ainsi présentés comme démonstration immédiate des préférences ou évitements matrimoniaux prévalant dans la société examinée. Or, les configurations ainsi décomptées sont inséparablement imbriquées avec des centaines de configurations différentes, dont les fréquences relatives ne sauraient être interprétées, telles quelles, comme des indicateurs comportementaux statistiquement significatifs. En effet,

l'idée naïve d'un lien simple et direct entre fréquence et préférence doit être abandonnée pour au moins deux raisons :

1. La fréquence d'un lien de parenté entre conjoints, aussi grande soit-elle relativement aux autres liens (de longueur et de structure comparables), n'est significative que si l'on connaît la fréquence *attendue* de cette configuration dans un réseau de parenté (1) où tous les mariages se seraient conclus au hasard et (2) qui présenterait exactement les mêmes lacunes et biais que le corpus généalogique considéré. Or, il n'y a aucun réseau réaliste où les mariages sont totalement indépendants de la parenté (personne n'épouse sa sœur, la co-résidence augmente à la fois la probabilité d'un mariage et d'un lien de cousinage, etc.), et les lacunes et biais d'un corpus sont eux-mêmes dépendants des structures de parenté en vigueur (une règle de résidence patrilocale augmente les lacunes dans le réseau des liens utérins, la polygamie introduit une distorsion agnatique, etc.). Pour estimer la significativité des motifs dans des réseaux de parenté, les méthodes classiques de statistique s'avèrent inadéquates – il faut une modélisation qui rende compte du caractère de réseau et des corrélations sous-jacentes entre les différentes observations. Certes, la modélisation statistique de réseaux sociaux a connu d'importantes évolutions depuis quelques années. Si la régression p1 permettait de modéliser les dyades (Holland et Leinhardt, 1981), elle ne prenait pas entièrement en compte les triades, et donc les aspects structurels du réseau (Federico de la Rúa, 2004). La régression p\* (Wassermann et Pattison, 1996) permet de lever ces contraintes, mais les résultats restent difficiles à interpréter. Le modèle Siena (Snijders, 2005) permet de modéliser des données de réseau dans un cadre longitudinal, mais dans un réseau complet. Toutefois, il n'existe pas encore d'études sur la transposition de ces méthodes dans le cadre des observations de motifs/structures, ni dans le cadre spécifique de réseaux de parenté (ou d'autres types de réseaux faiblement acycliques), dont le traitement peut exiger un autre cadre d'analyse.

2. Même si une configuration matrimoniale est reconnue comme significative, elle ne fait pas forcément l'objet, en soi, d'une préférence ou d'un évitement direct. La grande majorité des configurations matrimoniales qui se forment dans un réseau de parenté sont sans aucune signification sociologique intrinsèque, et représentent de simples *sous-produits* de comportements matrimoniaux visant à établir ou à éviter des configurations d'autres types. Cette interdépendance des anneaux matrimoniaux peut être d'ordre purement mathématique. Ainsi, par exemple, un évitement de mariages avec la sœur de la mère (MZ), combiné avec une préférence pour la fille de la sœur (ZD) se traduit automatiquement par l'évitement relatif de mariages avec la fille de la sœur du père (FZD), etc. (Leach, [1952] 1968 ; Rivière, 1969). Mais cette interdépendance peut aussi avoir des soubassements sociologiques. Chez les Peul (Sénégal) par exemple, le fils peut d'autant plus facilement épouser la fille d'un frère de mère (MBD) si son père avait lui-même épousé une fille de frère de père (FBD) (Barry, 1998).

Plus généralement, la détermination empirique des relations dans les réseaux empiriques peut souffrir de divers biais génériques (Stork et Richards, 1992 ; Little et Schenker, 1995 ; Robins *et al.*, 2004 ; Kossinets, 2006) : dans le cas le plus simple, le contexte ne permet pas de

collecter de données sur la totalité des acteurs, mais seulement sur une partie d'entre eux. Tandis que ce type de biais est généralement sans conséquence importante dans l'étude traditionnelle des populations, où un plus petit échantillon d'acteurs n'entraîne qu'une moindre représentativité et significativité statistique des résultats, il en va tout autrement dans les études de réseaux où la définition de la « frontière » influence fondamentalement les résultats (Laumann *et al.*, 1989). L'absence d'une portion même limitée du réseau peut effectivement modifier radicalement les propriétés *topologiques* de celui-ci, parce que la plupart de ces propriétés sont définies de manière *holiste*. Ce phénomène est déjà bien connu par exemple dans le cas de la percolation, où la connexité n'est soudainement plus assurée dès lors que l'on dépasse un seuil critique sur la densité des liens (Bollobas, 1985). L'étude des renchaînements en tant que 2-connexité du réseau de parenté est ainsi particulièrement concernée par ce type de biais ; où l'absence d'un seul lien empêche de fermer un cycle tout entier.

Les problèmes évoqués – les biais du corpus et l'interdépendance des configurations matrimoniales – ont jusqu'à présent rendu très hasardeuse toute interprétation du profil matrimonial d'un réseau de parenté empirique, sauf dans le cas (et dans la mesure) où ils confirment une règle matrimoniale connue à l'avance. Pour pouvoir évaluer l'impact des règles matrimoniales sur la structure des réseaux de parenté, et, réciproquement, pour pouvoir tirer des inférences sur la logique des comportements matrimoniaux à partir d'une analyse des réseaux empiriques de parenté, il faut avant tout pouvoir comprendre la dynamique interne des réseaux de parenté, leur articulation avec des facteurs non généalogiques, et leur effet sur les conditions de leur propre observation.

Les méthodes exigées par une telle recherche s'inscrivent dans le champ de la simulation des réseaux sociaux. Les caractéristiques mathématiques spécifiques des réseaux de parenté constituent toutefois un contexte de modélisation particulier. D'une part, il s'agit de modéliser la distribution et la composition de cycles dans des graphes faiblement acycliques à partir d'une règle portant sur la formation (ou non) d'une classe de cycles dûment définie ; d'autre part, il s'agit de modéliser la reconstruction d'un graphe à partir d'un ensemble de sous-graphes égocentrés dont la taille et la structure dépendent de la position d'Ego.

L'idée de tester la significativité des circuits matrimoniaux par une simulation sous l'hypothèse d'un choix matrimonial aléatoire remonte à Hammel (1976), qui l'a utilisée comme outil critique contre les modèles mécaniques de la théorie structuraliste de l'alliance. Depuis cette date, plusieurs logiciels ont été développés pour simuler les comportements matrimoniaux sous certains paramètres et contraintes à la fois démographiques et culturels (voir par exemple Lang, 1995). Certains de ces logiciels – comme *GeneoRnd* de V. Batagelj (<http://vlado.fmf.uni-lj.si/pub/networks/pajek/howto/geneoRnd.htm>) – permettent la simulation directe de réseaux généalogiques. En revanche, ces modèles ne permettent pas de modéliser l'interdépendance des circuits matrimoniaux ou les impacts des biais mnémoniques. De premiers pas dans ce sens ont été récemment amorcés. Ainsi, la méthode de White (1999) de permutation aléatoire de réseaux empiriques ([http://intersci.ss.uci.edu/wiki/index.php/Software:\\_Kinship\\_simulation](http://intersci.ss.uci.edu/wiki/index.php/Software:_Kinship_simulation)), permet de conserver non seulement des

caractéristiques démographiques, mais aussi certains des biais mnémoniques du réseau de départ. Le modèle de Read (1998) simule l'impact de règles matrimoniales, formulées en termes du système culturel (en l'occurrence !kung san), sur l'évolution démographique et les alliances entre villages. Toutefois, il n'existe pas encore de méthodes pour simuler, de façon générale, l'impact des règles matrimoniales sur la distribution globale des circuits matrimoniaux dans un réseau de parenté, ni pour modéliser explicitement les biais mnémoniques qui caractérisent les corpus généalogiques.

L'enjeu du projet consiste donc, à modéliser l'évolution de ces réseaux faiblement acycliques que sont les réseaux de parenté, aussi bien au niveau « réel » (i.e. des liens de parenté existant effectivement), par le biais des comportements (matrimoniaux) qu'au niveau « virtuel » (i.e. des liens dont se souviennent les acteurs) par le biais de la mémorisation (généalogique), ces deux niveaux étant étroitement interconnectés.

## **2.2. OBJECTIFS ET CARACTERE AMBITIEUX/NOVATEUR DU PROJET**

L'objectif de SimPa est d'aboutir à un modèle intégré de l'émergence de circuits matrimoniaux dans les réseaux de parenté, ainsi qu'à des techniques de simulation permettant la modélisation réaliste des réseaux utilisés par les chercheurs (anthropologues, démographes, historiens et sociologues) étudiant ces réseaux. Ces techniques seront implémentées dans des logiciels librement accessibles et facilement utilisables. En même temps, ce projet contribuera, par l'analyse de la dynamique d'un type spécifique de réseau, à la compréhension de la morphogenèse des réseaux faiblement acycliques et à la théorie statistique d'événements interdépendants. Cette double ambition – fournir aux sciences sociales un instrument d'analyse indispensable et faire avancer un domaine méthodologique important de la mathématique des réseaux – se reflète dans la composition des porteurs du projet, qui réunit mathématiciens, statisticiens, démographes et anthropologues.

Plutôt que de simplement associer « fournisseurs » et « utilisateurs » de modèles de simulation, le projet est motivé par la nécessité de réunir des compétences différentes dans le développement de produits finaux qui profiteront également aux disciplines impliquées : d'une part, des algorithmes et logiciels de simulation applicables aux données traitées par les anthropologues et les démographes, d'autre part, la formalisation d'une structure mathématique spécifique qui peut servir de modèle pour l'avancement de la théorie des graphes faiblement acycliques, dont le champ d'application s'étend bien au-delà du domaine de la parenté (réseaux de citations, etc.).

Le but immédiat du projet consiste à résoudre, par voie de simulation contrôlée, une série de problèmes qui entravent jusqu'à présent l'interprétation de recensements de circuits dans des réseaux de parenté, et, partant, toute tentative de fonder l'analyse d'un système de parenté sur l'examen des pratiques matrimoniales empiriques : ces problèmes concernent la **distribution attendue de circuits** sous des hypothèses réalistes, la neutralisation des **biais du corpus**, et l'identification des **artefacts du réseau**.

Il s'agit donc de développer des modèles de l'évolution matrimoniale et mnémonique des réseaux de parenté, d'implémenter ces modèles dans des algorithmes et des logiciels facilement maniables et librement accessibles par les chercheurs en sciences sociales, et de les tester sur un échantillon d'une vingtaine de réseaux de parenté empiriques (issus de sociétés contemporaines et historiques, « modernes » et « traditionnelles »). Les modèles développés doivent être théoriquement neutres, c'est-à-dire susceptibles d'être paramétrés selon les approches théoriques des chercheurs, sans leur imposer a priori des hypothèses sur le fonctionnement des comportements matrimoniaux ou sur les mécanismes de mémorisation.

Puisqu'il s'agit de modéliser la *totalité* des configurations matrimoniales interdépendantes, sans restriction préalable sur des motifs censés exprimer une règle matrimoniale, la démarche préconisée consiste à se concentrer sur les interactions entre un type de configurations jusqu'à présent privilégié dans les études de parenté – les mariages consanguins – et un type de configurations très peu étudié en dépit de son importance sociologique – les renchaînements de mariage (mariages entre affins). Le fait que ces derniers, impliquant plusieurs liens de mariage, ne fassent pas toujours l'objet de règles matrimoniales explicites, les qualifie en effet comme un cas test pour étudier les effets des règles matrimoniales sur la distribution globale des circuits matrimoniaux dans un réseau de parenté.

Le traitement de ces problèmes promet des connaissances approfondies sur la topologie des réseaux faiblement acycliques, dont les réseaux de parenté ne sont qu'un cas particulier. En effet, les réseaux de parenté offrent la possibilité de résoudre plusieurs problématiques théoriques concernant les approximations présentes dans des données empiriques dans un cas où les règles de relation et de formation semblent *a priori* plus strictes, plus contraignantes et donc aussi plus *régulières* que pour de nombreux autres réseaux d'interaction :

- (1) Les empêchements sociologiques (comme l'inceste) ou structurels (comme ceux liés à la chronologie) qui contraignent formellement l'existence de certains motifs dans les réseaux de parenté semblent beaucoup moins présents dans d'autres réseaux sociaux réels (comme par exemple les réseaux de collaboration, ou de communautés de blogueurs) où les régularités strictement liées aux biais mentionnés ci-dessus semblent ainsi davantage bruitées.
- (2) La signification du lien est très rigide : « être l'enfant de A » semble être moins polysémique que « connaître B », « être l'ami de C », etc. Les réseaux de parenté semblent ainsi présenter moins de difficultés de définition que d'autres réseaux (« avoir interagi avec quelqu'un » – mais comment ?, « avoir participé au même projet » - mais est-ce interagir forcément ?, « connaître quelqu'un » ou « en être l'ami » - à quel point et de quelle manière?, etc.).

Le projet manifeste donc à la fois un intérêt pratique éminent pour les études de parenté, en ce qu'il promet de débloquent la voie pour l'identification et l'interprétation des motifs

significatifs dans des réseaux de parenté, et un intérêt théorique plus large pour la théorie des graphes et la méthodologie de simulation. A ce dernier égard, le projet représente une étude pionnière : il n'existe pas, jusqu'à présent, un modèle intégré de simulation sous contrainte de données manquantes ou capable de contrôler les artefacts du réseau, et le traitement statistique de motifs réticulaires (en tant qu'événements interdépendants) est toujours à l'état embryonnaire.

La contribution du projet SimPa relève donc aussi bien des sciences formelles (mathématique et statistique) que des sciences sociales (études de parenté). En fait, les problèmes principaux à traiter – telle l'interdépendance entre circuits matrimoniaux d'une part, entre position généalogique et mémoire généalogique d'autre part – se présentent à la fois sous un aspect mathématique et sociologique, et un des grands défis de la théorie des réseaux sociaux consiste à pouvoir distinguer les effets des deux types de déterminations. Sans cela, on risque fort de vouloir chercher des interprétations sociologiques pour des faits qui sont purement mathématiques.

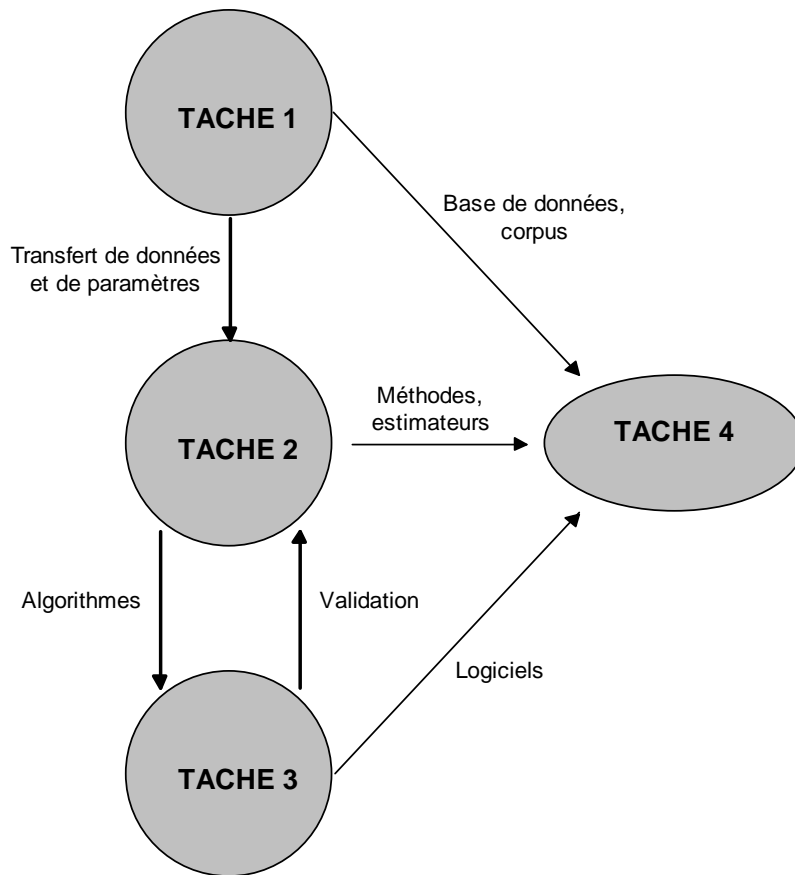
Les indicateurs de réussite du projet, pouvant servir de critère d'évaluation, se rapportent à ce double objectif : fournir aux études de parenté des instruments pertinents pour l'interprétation des données, et développer des concepts mathématiques et statistiques d'application plus large dans l'analyse des réseaux sociaux.

### 3. PROGRAMME SCIENTIFIQUE ET TECHNIQUE, ORGANISATION DU PROJET

#### 3.1. PROGRAMME SCIENTIFIQUE ET STRUCTURATION DU PROJET

Le projet s'organise en quatre grandes tâches : la **collecte** et l'**analyse des données** (tâche 1) servira de base pour la **modélisation** et la **simulation** des réseaux de parenté (tâches 2 et 3) en fournissant les paramètres de modélisation et les réseaux de référence pour la simulation. Les deux tâches centrales de la modélisation et de la simulation procéderont en communication réciproque, les modèles étant implémentés en algorithmes et soumis à la validation par simulation. Ces trois tâches produiront des résultats (corpus généalogiques, méthodes et estimateurs, logiciels) dont la **valorisation** et la **diffusion** (en ligne, dans des publications, et dans le cadre d'ateliers) constituera la tâche 4 du projet.

La figure 1 fournit les principaux liens qu'entretiennent les tâches entre elles. Le tableau 1 récapitule les implications en personnes-mois des différents partenaires sur les différentes tâches, ainsi que l'échéancier des jalons et livrables correspondant à chaque tâche.



**Figure 1 : Organigramme technique**

	Partenaires (Souligné = Responsable)			Chronogramme / Chemin critique						Total
	INED	CAMS	CEMAF	Année 1		Année 2		Année 3		
				M6	M12	M18	M24	M30	M36	
Tâche 0	<u>9</u>	0	0	1.5	1.5	1.5	1.5	1.5	1.5	<b>9</b>
Tâche 1	4.8	0	<u>28.2</u>	7	7	7	6	6	0	<b>33</b>
Tâche 2	<u>24</u>	16.2	15.8	9	9	10	10	10	8	<b>56</b>
Tâche 3	0	<u>16.2</u>	5.8	0	4	4	5	5	4	<b>22</b>
Tâche 4	<u>6</u>	0	0	0	0	0	2	2	2	<b>6</b>
<b>Total</b>	<b>43.8</b>	<b>32.4</b>	<b>49.8</b>	<b>17.5</b>	<b>21.5</b>	<b>22.5</b>	<b>24.5</b>	<b>24.5</b>	<b>15.5</b>	<b>126</b>

Livrables / Jalons	L-0.1 L-1.1	L-1.2 L-4.1	L-1.3 J-4.2	L-1.4	L-4.4 J-3.1	J-4.3 L-2* L-3.1 L-3.2	
Rapports d'avancement + état des dépenses		x		x		x	

**Tableau 1 : Répartition des personnes.mois en fonction des tâches et des partenaires**

La **tâche 0**, dont l'INED est le responsable, consiste à **organiser** et à **synchroniser les activités des partenaires**, à assurer la **gestion administrative** du projet, et à garantir la **visibilité publique** de son avancement (voir la section 3.2 ci-dessous).

La **tâche 1**, coordonnée par le CEMAF, consiste à coordonner la **collecte**, le **codage** et l'**analyse** des matériaux généalogiques qui sont le support des tâches de modélisation et de validation.

La **tâche 2**, tâche de **modélisation**, sera coordonnée par l'INED. Au sein de cette tâche, le CEMAF prendra en charge l'analyse comparative de la distribution de circuits dans les réseaux empiriques (sous-tâche 2.1), le CAMS s'occupera de la modélisation de l'interaction et de la composition de circuits au sein des réseaux de parenté (sous-tâche 2.2), et l'INED sera responsable pour le développement des estimateurs de probabilité, des modèles de distribution et des mesures de significativité statistique pour les circuits matrimoniaux (sous-tâche 2.3).

La **tâche 3**, coordonnée par le CAMS et étroitement liée à la tâche 2, consiste à développer les algorithmes de **simulation** afin de valider les modèles développés dans le cadre de la tâche 2.

Enfin, l'INED en tant que coordinateur se chargera de la **tâche 4**, c'est-à-dire la **valorisation** et **diffusion** des résultats des tâches 1-3.

### **3.2. MANAGEMENT DU PROJET**

**Voir aussi section 4 ci-dessous**

Le coordinateur du projet se chargera d'organiser mensuellement une réunion de tous les partenaires.

Un site web collaboratif sera mis en place dès le premier mois afin de permettre la synchronisation entre partenaires, et notamment, par le biais d'une partie publique, mettre en valeur les différentes avancées des tâches qui suivent. Enfin, un colloque conclusif sera organisé à la fin du projet afin d'assurer une large publicité des résultats du projet.

Le coordinateur sera assisté par un administrateur (embauché sur un CDD à quart-temps dans le cadre de l'INED).

Livrable L-0.1. Site web du projet

Livrable L-0.2. Rapports semestriels

### 3.3. DESCRIPTION DES TRAVAUX PAR TACHE

#### 3.3.1 TACHE 1 – COLLECTE ET CODAGE DES DONNEES [RESPONSABLE : CEMAF, INED]

Les matériaux généalogiques qui serviront de support pour le développement et la validation des modèles (tâches 2 et 3) se divisent en deux groupes :

1. d'une part, des corpus relativement complets et peu biaisés qui peuvent être considérés, dans certaines limites, comme une représentation fiable du réseau généalogique réel de la société étudiée. Il s'agit ici surtout de données dont la collecte a pu s'appuyer sur des sources écrites bien documentés (sociétés européennes à partir de l'introduction de l'Etat civil), ou fondées sur des recherches de terrain dans des communautés relativement restreintes, endogames et/ou sans règle résidentielle unilocale. Parmi les corpus de ce type déjà publiés et codés figurent par exemple les généalogies de la population de San Marino (Gasperoni, 2008) ou celles des Touareg Iwellemmeden Kel-Kummer du Mali (Chaventré, 1983).

2. d'autre part, des corpus manifestant des lacunes et des biais, dus à des mécanismes inhérents à la société étudiée qui sont relativement bien connus et documentés (règles de résidence, systèmes de filiation, exogamie, rites funéraires et matrimoniaux, organisation politique, etc.). Tout en représentant une image déformée de la société, ces matériaux permettent d'identifier les causes de ces distorsions, et serviront à étudier les interactions entre mémoire généalogique et structure de parenté. Parmi les corpus déjà codés de cette catégorie figurent par exemple les généalogies des communautés juives italiennes du 16<sup>e</sup> siècle (Gasperoni, 2006) ou les Ouatchi du Togo de Sud-Est (Hamberger, 2009). On analysera dans ce cadre aussi le corpus collecté chez les Bwa par Véronique Hertrich dans le cadre du projet INED « Dynamique démographique, conjugalité et parentalité en Afrique », corpus dont la structure patrilinéaire pourra éventuellement être contrastée avec les résultats d'un recueil complémentaire de données, permettant ainsi d'étudier l'évolution du biais agnatique dans le cadre d'une expérience contrôlée sur le terrain.

**Sous-tâche 1.1 : Codage et saisie** de corpus publiés (dans des ouvrages ou sur le web), ou obtenus d'autres chercheurs, de la catégorie 1 (env. 10 corpus). [Responsable : CEMAF, INED]

Les corpus seront codés et analysés en format .tip à l'aide du logiciel Puck. Cette sous-tâche peut être accomplie par des vacataires ayant reçus une formation de base.

**Sous-tâche 1.2 : Complétion et équilibrage**, au moyen de travaux d'archive ou d'enquêtes de terrain, de corpus existants mais lacunaires ou biaisés (5-10 corpus) [Responsable : CEMAF].

**Sous-tâche 1.3 : Enquêtes** systématiques sur la **mémoire généalogique** individuelle et collective dans des sociétés pour lesquelles des corpus généalogiques existent déjà (5 corpus) [Responsable : CEMAF, INED]. Cette sous-tâche exigera une réflexion méthodologique préalable.

On prendra soin de garder un équilibre entre des corpus s'appuyant sur des sources écrites et orales, et relevant des sociétés contemporains et historiques, « modernes » et « traditionnelles ».

Les corpus informatisés et anonymisés seront rendus publiquement accessibles en ligne (L-4.1).

<u>Livrable L-1.1.</u>	10 Corpus publiés codés
<u>Livrable L-1.2.</u>	10 Corpus compétés codés
<u>Livrable L-1.3.</u>	Analyse synoptique des corpus
<u>Livrable L-1.4.</u>	Rapports des enquêtes de mémoire généalogique

### 3.3.2 TACHE 2 – MODELISATION [RESPONSABLE : INED]

La tâche de modélisation de la dynamique matrimoniale et mnémonique sous-jacente à la distribution de circuits dans les réseaux de parenté empiriques sera abordée par plusieurs voies, impliquant différentes contributions des partenaires :

**Sous-tâche 2.1. Analyse comparative** de distributions de circuits dans des réseaux empiriques. [Responsable : CEMAF]

La première méthode consiste à considérer les fréquences relatives d'un nombre restreint de types de circuits matrimoniaux (par exemple les mariages entre cousins de 2<sup>e</sup> degré ou les mariages entre affins proches), afin d'identifier des profils inattendus et de les mettre en relation avec (a) les règles et prohibitions matrimoniales connues, (b) d'autres facteurs sociologiques, par exemples résidentiels, (c) des effets démographiques (taux de polygamie ou de remariage, nombre d'enfants, etc.), (d) des biais de mémorisation ou d'enquête, et (f) des effets indirects résultant de la combinaison de ces différents facteurs (pour un exemple d'une analyse de ce type voir Daillant et Hamberger, 2009).

Ces analyses, menées, si possible, en association avec les chercheurs ayant recueilli les données, pourront s'appuyer sur une bonne connaissance du contexte sociologique, nécessaire pour pouvoir distinguer entre les différentes catégories d'effets évoqués. En particulier, elles se fonderont sur les enquêtes de mémoire généalogique fournies dans le cadre de la tâche 1.3.

L'instrumentaire méthodologique de ces analyses consistera en des outils relevant de l'analyse de réseaux (comme le réseau d'intersections de circuits), en des instruments d'ordre algébrique (comme la table de compositions de circuits) et en des mesures statistiques pour évaluer la significativité et la probabilité conditionnée par les fréquences relatives de circuits matrimoniaux. Il s'agit donc d'instruments qui, bien qu'en partie déjà implémentés dans des

outils informatiques disponibles, seront continuellement produits et élaborés dans le cadre des sous-tâches 2.2 et 2.3.

**Sous-tâche 2.2.** Modélisation de l'interaction et de la **composition de circuits** au sein de réseaux de parenté [Responsable : CAMS]

La composition des circuits matrimoniaux dans la formation de circuits plus complexes ou de structures plus larges est un des sujets les plus anciens de la mathématique de la parenté. Il a été longtemps traité par la construction de « modèles mécaniques » de systèmes de parenté (surtout de structures « élémentaires »), c'est-à-dire des réseaux de parenté où chaque individu conclut exactement le même type de mariage (par exemple, quand tout le monde épouse la cousine croisée matrilatérale MBD, celle-ci sera toujours forcément aussi une MMBDD). Ces modèles, souvent appuyés sur la théorie de groupes de permutation, ont ensuite été étendus aux types de mariage plus complexes (voir Tjon Sie Fat 1990). Toutefois, il n'existe jusqu'à présent pas de méthode pour traiter de façon générale la composition de circuits matrimoniaux, ni d'élargir cette analyse d'un cadre strictement mécanique à un cadre statistique où la fréquence d'un circuit d'un type donné serait fonction de la *probabilité d'intersection* de deux circuits de types différents.

Cette tâche est préliminaire à toute interprétation de motifs apparaissant dans des réseaux réels ou simulés. En effet, la combinaison de circuits dans un réseau de parenté conduit toujours à l'apparition *mécanique* de certains motifs qui, plutôt que d'être représentatifs des pratiques interactionnelles et matrimoniales des acteurs, ne sont que des artefacts. Ce phénomène est bien connu dans la théorie des réseaux sociaux. Par exemple, les réseaux d'interaction construits sur la base de participation conjointe à des événements (et donc modélisables par des graphes bipartis) sont amenés à exhiber naturellement cliques, transitivité et fort *clustering* (Newman *et al.*, 2002).

Cette thématique recoupe l'étude des réseaux aléatoires qui constituent des modèles de référence (« *null models* ») dans la reconnaissance de structures atypiques dans les réseaux empiriques. Dans l'ensemble, ces modèles s'appuient :

- Soit sur des paramètres plus ou moins riches dont les valeurs sont fixées par le réseau empirique, afin de vérifier si certaines caractéristiques stylisées suffisent à expliquer l'existence ou non de certains motifs. C'est le cas par exemple des modèles de configuration (Newman *et al.*, 2002) où la distribution complète des degrés du réseau d'origine est utilisée comme unique paramètre du réseau aléatoire reconstruit. Ce type de modèle se prête toutefois assez peu à l'explication et à l'interprétation des mécanismes menant à la structure réelle observée.
- Soit sur des processus morphogénétiques fondés sur des mécanismes stylisés de croissance et d'évolution du réseau (Barabasi et Albert, 2002). Généralement, les hypothèses sont choisies afin d'être les plus génériques et peuvent ainsi souvent sembler relativement déconnectées de l'intuition anthropologique.

Dans tous les cas, cette approche vise à expliquer l'existence de motifs empiriques par le biais des seules hypothèses qui fondent le modèle ; ce type de reconstruction fournit ainsi un modèle de référence qui permet d'expliquer immédiatement la présence de motifs empiriques comme des phénomènes directement engendrés par construction du modèle.

L'analyse de cette interdépendance des circuits dans les réseaux de parenté a progressé au fil des années par le développement (et la mise en œuvre informatique) d'outils conceptuels comme la représentation des circuits par des vecteurs caractéristiques (facilitant l'étude de leur structure algébrique) et la mise en évidence de leur interconnexion empirique au moyen de réseaux de deuxième ordre où les sommets représentent des types de circuits et les lignes leurs intersections, c'est-à-dire les mariages qui intègrent simultanément des circuits des deux classes. Ces réseaux de deuxième ordre, qui peuvent être soumis à l'instrumentaire classique de la théorie des réseaux (densité, centralité, recensement triadique, etc.), mais aussi à des analyses spécifiques telle que la méthode de décomposition contrôlée, représentent un premier outil pour accéder à la topologie des circuits matrimoniaux. Il doit être développé davantage afin d'aboutir à un outil plus élaboré, tel un treillis complet représentant les relations d'intersection et d'inclusion de circuits. Ces outils de recherche empiriques (voir par exemple Daillant et Hamberger, 2009 ou Gabail et Kyburz, 2009) sont complétés par des outils mathématiques qui relèvent de l'algèbre des circuits matrimoniaux. L'application de ces outils d'analyse à des réseaux à la fois empiriques et simulés, doit permettre de comprendre la *corrélation* mutuelle des fréquences de circuits matrimoniaux appartenant à des types différents qui, jusqu'à présent, ont été traités à tort comme indépendants.

Nous nous placerons ici dans la classe des modèles s'appuyant à la fois sur des mécanismes réalistes, i.e. empiriquement validés, et exprimables simplement sous forme de contraintes algébriques, c'est-à-dire à même de présenter une interprétation anthropologique immédiate. Plus simplement, il s'agira donc de vérifier si certaines règles matrimoniales et contraintes algébriques de construction des réseaux de parenté sont à même d'expliquer certaines configurations, en simulant des réseaux aléatoires qui respectent ces contraintes connues.

La tâche consiste donc à développer, premièrement, un instrumentaire conceptuel et méthodologique adéquat à l'algèbre des circuits matrimoniaux, ainsi que des outils pour représenter et analyser la topologie de ces circuits (intersections, inclusions, enchaînements, etc.) par des structures graphiques telles que des graphes ou des treillis. Deuxièmement, il s'agira d'élaborer des outils de simulation contrôlée pour appliquer cet instrumentaire dans un contexte probabiliste, et développer des critères pour décider si des configurations apparemment remarquables dans un réseau de parenté simulé sont simplement fortuites (artefacts du réseau) ou bien véritablement typiques de certaines formes de contraintes et de pratiques matrimoniales.

Plus largement, il s'agira de mener une opération duale de celle, classique, de caractérisation statistique de la dynamique d'un réseau empirique: en effet, tandis que les outils existants

s'attachent à démêler les influences conjointes que certaines configurations structurelles peuvent avoir sur la probabilité d'apparition des liens ou dyades (Snijders, 2001), nous développerons ici un cadre statistique original visant à estimer les influences conjointes que certains liens ont sur l'existence ou non des divers motifs structurels. En d'autres termes, les méthodes existantes estiment la contribution respective de divers motifs dans l'apparition de certains liens. Nous visons à estimer la contribution respective des divers mécanismes de création de liens sur l'apparition de certaines structures. Concrètement, il s'agit donc de voir si certaines structures sont plus ou moins naturellement déterminées par certains mécanismes de création de liens, tout en prenant en compte les corrélations possibles entre structures – c'est notamment le cas lorsque certains motifs induisent systématiquement d'autres motifs (par exemple, l'existence d'une clique de taille 4 induit nécessairement l'existence de quatre cliques de taille 3).

**Sous-tâche 2.3.** Développement des **estimateurs de probabilité**, des **modèles de distribution** et des **mesures de significativité** statistique pour les circuits matrimoniaux [Responsable : INED]

Avant de juger si la fréquence d'un circuit matrimonial est révélatrice d'une préférence ou d'un évitement matrimonial ou d'une dynamique sociologique privilégiant ou défavorisant un type de parents comme conjoints potentiels, il est nécessaire de savoir si cette fréquence est statistiquement significative. Une approche naïve de cette question consisterait à comparer la fréquence relative des chaînes de parenté d'un certain type (par exemple entre cousins croisés) avec la fréquence relative des circuits matrimoniaux (donc des chaînes closes) correspondantes, afin de repérer le caractère sur- ou sous-proportionné des circuits en question. Or, cette approche, qui est adaptée au traitement d'un échantillon d'événements *indépendants*, n'est pas appropriée dans un contexte réticulaire où tous les événements – c'est à dire formations de circuits – sont presque par définition dépendants les uns des autres, et ce, aussi bien mathématiquement que sociologiquement. Cette interdépendance oblige à repenser le calcul des probabilités d'apparition d'un circuit matrimonial au sein des chaînes de parenté de même type.

Il est donc une exigence de premier ordre pour l'analyse des structures matrimoniales de développer des estimateurs de la probabilité de formations cycliques dans un réseau. Ces estimateurs devront se fonder sur un certain nombre de mesures démographiques standard à déterminer (nombre d'enfants, de conjoints, etc.), et des mesures classiques pour l'analyse de réseaux, dont certains (comme la densité, la centralité ou la connectivité) exigent une adaptation pour le cas spécifique des réseaux de parenté (par exemple, la densité matrimoniale est une fonction négative de la profondeur du réseau, puisque la probabilité d'un mariage diminue forcément avec la distance générationnelle). En effet, ces mesures classiques dépendent de la topologie du réseau et des conséquences des interdits sociologiques sur cette topologie. Le développement de ces mesures pourra s'appuyer sur les corpus établis dans le cadre de la tâche 1 et sera directement lié aux résultats des simulations entreprises dans le cadre de la tâche 4. Le projet s'attaquera ainsi au problème de la représentativité de motifs dans les graphes, problème sur lequel la littérature se contente

généralement de « z-score » (cf., par exemple, Milo *et al.*, 2004) sans prendre en compte l'effet des corrélations entre différents motifs.

L'un des intérêts parmi d'autres d'une telle analyse est de pouvoir identifier des motifs dont l'occurrence dans un réseau purement aléatoire serait beaucoup moins fréquente, et qui représentent donc une caractéristique significative du réseau observé par rapport à un réseau simulé comparable. Les problèmes d'inférence à partir d'un corpus donné, visant à établir des règles pour la population dont il est issu, ont fait l'objet de peu de travaux jusqu'à présent. On étudiera dans ce cadre les diverses tentatives de modélisation déployées en analyse de réseaux, telles les régressions  $p_1$  (Holland et Leinhardt, 1981), random graph models ou  $p^*$  (Wassermann et Pattison, 1996), adaptées au cadre spécifique des réseaux de parenté. On étudiera aussi, comme il a été précisé au point précédent (sous-tâche 2.2), comment adapter une méthodologie du type utilisé par Snijders à un cadre conceptuel inversé.

Le problème est accentué par le fait que dans ce type de réseau, les données ou types de liens peuvent être très souvent lacunaires, et que ces lacunes ne sont pas répercutées au hasard. Il est donc extrêmement important de posséder des mesures pour évaluer et rendre comparables ces biais et lacunes, et des techniques permettant d'ajuster les estimateurs de significativité en connaissance de ces biais. On développera une série de mesures pour décrire les biais dans les corpus utilisés dans le cadre de ce projet, pour lesquels le contexte de collecte et la nature des lacunes sont particulièrement bien connus. On listera par exemple un certain nombre d'indicateurs permettant de jauger du déséquilibre hommes-femmes ou agnatique-utérin au sein d'un tel réseau, afin d'estimer s'il existe un déséquilibre des genres au niveau des sources. On étudiera successivement des indicateurs verticaux permettant de juger de la profondeur d'un corpus, et des indicateurs horizontaux permettant de déterminer si les données lacunaires sont liées.

<u>Livrable L-2.1</u>	Rapport sur l'analyse comparative des structures matrimoniales dans des corpus empiriques
<u>Livrable L-2.2</u>	Rapport sur les outils conceptuels pour l'analyse des algèbre et topologie de circuits matrimoniaux
<u>Livrable L-2.3</u>	Rapport sur les mesures et estimateurs statistiques de la distribution des circuits matrimoniaux

### 3.3.3 TACHE 3 – SIMULATION ET VALIDATION [RESPONSABLE : CAMS]

La tâche de la simulation des réseaux de parenté sera effectuée en parallèle et en communication continue avec la tâche de modélisation, l'une fournissant des inputs pour l'autre et vice versa. La simulation permettra ainsi de valider et/ou rectifier les modèles développés en comparant leurs résultats avec les corpus de données empiriques, tandis que la modélisation à partir de l'analyse des corpus réels permettra de simuler les corpus artificiels qui serviront d'étalon pour l'interprétation des premiers.

L'interaction entre corpus empiriques (des deux catégories décrites dans la section 3.3.1 : complets ou incomplets), modélisation et simulation de corpus artificiels prendra deux formes différentes :

**Sous-tâche 3.1.** Simulation, au moyen d'une modélisation des comportements matrimoniaux, de généalogies **complètes** à partir de paramètres comparables avec ceux de généalogies **réelles, complètes ou incomplètes**.

La collecte souffre généralement de divers biais de construction : nous l'avons souligné précédemment, certains liens et acteurs vont être plus fréquemment ou facilement mentionnés que d'autres, suivant des processus matériels et/ou sociologiques visant à surestimer et privilégier certains éléments du réseau. Dans cette optique, il devient crucial de déterminer (i) si le réseau effectivement collecté est représentatif du réseau complet sous-jacent supposé (manque-t-il des liens, des acteurs, de quelle manière ? certains chemins de construction du graphe ont-ils privilégié certains acteurs ?) et, le cas échéant, (ii) comment appréhender ces réseaux partiels, pour lesquels les mécanismes à l'origine des biais constatés ou supposés ont été mis en évidence par d'autres études ethnographiques.

La littérature existante sur le problème des réseaux incomplets, relativement peu fournie (Robins *et al.*, 2004), souligne principalement l'effet des données manquantes sur les observations empiriques. Par exemple, la fameuse étude de Travers et Milgram (1969), qui a visé à estimer la distance moyenne entre deux individus de la population étatsunienne par l'intermédiaire d'une expérience fondée sur des chaînes de lettres, a été réexaminée récemment par Dodds *et al.* (2003) : puisque, dans l'étude originale, la plupart de ces lettres ne sont pas arrivées à destination, cette étude reproduit l'expérience sur un réseau de contacts par e-mail en s'attachant à estimer l'effet des non-réponses sur l'évaluation de cette distance moyenne. D'autres travaux encore montrent l'inobservabilité de certains effets présents sur des réseaux de grandes tailles lorsque des échantillons de petite taille sont utilisés (Dorogovtsev *et al.*, 2001).

Dans la plupart des cas cependant, ces études discutent de l'existence de biais et de leur effet sur les mesures, sans pour autant proposer de solution de reconstruction du graphe empirique. En d'autres termes, *il s'agit dans ces travaux de réparer les mesures plutôt que le graphe lui-même* ; en décrivant une incertitude sur les mesures (Dodds *et al.*, 2003) ou bien en utilisant les données réelles pour en extrapoler les mesures qui auraient dû être observées si l'on avait disposé du réseau exhaustif réel sous-jacent (Robins *et al.*, 2004 ; Kossinets, 2006). Des travaux préliminaires visant à reconstruire certaines parties du réseau à l'aide d'hypothèses très simplistes liées à la symétrie des liens ont été proposés par Stork et Richards (1992). Nous nous plaçons dans ce dernier cadre en proposant d'étendre la reconstruction à des classes générales de biais de collecte (et non pas seulement la symétrisation des liens).

L'objectif de cette sous-tâche sera donc de proposer des procédures de complétion de graphe s'appuyant sur des processus connus, dont le résultat sera validé par la structure de réseaux supposés exhaustifs.

**Sous-tâche 3.2.** Simulation, au moyen d'une modélisation de la mémoire généalogique, de généalogies **incomplètes** engendrées à partir de généalogies **complètes, réelles ou simulées** ; et simulation, au moyen d'une modélisation des comportements matrimoniaux, de généalogies **incomplètes** engendrées à partir de généalogies **incomplètes réelles**.

Les techniques habituellement utilisées pour la simulation des réseaux de parenté visent à produire un réseau comparable aux réseaux de parenté réels. Or, les réseaux dont disposent les chercheurs en sciences sociales sont tous des réseaux *reconstruits à partir des sources* (orales ou écrits) dont les lacunes et biais sont directement corrélés avec la position des informateurs dans le réseau de parenté concerné. A ce biais de premier ordre s'ajoute un biais de deuxième ordre introduit par le choix des réseaux égocentrés utilisés comme base pour la reconstruction du réseau. Ce choix à son tour (et bien que déterminé par le chercheur) est souvent soumis à des contraintes relevant du réseau de parenté lui-même. Les conditions de reconstruction du réseau sous ce double biais doivent faire l'objet d'une modélisation en soi, non seulement afin de pouvoir neutraliser ces biais et faire des inférences sur le réseau « réel » sous-jacent, mais aussi afin de comprendre la reproduction des réseaux de parenté dans le souvenir des intéressés, étant donné que ce sont précisément les réseaux reconstruits par les divers acteurs qui déterminent les choix matrimoniaux effectifs.

Plusieurs techniques ont été développées pour simuler à la fois le réseau de parenté et sa reconstruction en un corpus généalogique. Une méthode (développée par White, 1999) consiste à partir de corpus généalogiques empiriques et de les *permuter* sous certaines conditions (telle l'invariance d'une série de paramètres démographiques choisis) afin d'obtenir un réseau « aléatoire » produit sous les mêmes contraintes que le réseau original, et donc comparable avec celui-ci. Une approche alternative (implémentée dans le logiciel *Puck*) consiste à modéliser explicitement la collecte des données généalogiques (*virtual fieldwork*). Les deux méthodes ne sont pas encore parvenues à surmonter les difficultés posées par l'interaction complexe entre choix matrimoniaux et mémoire généalogique : la méthode de permutation des partenaires de mariage détruit inévitablement une série de caractéristiques structurelles des réseaux reconstruits (distribution inégale des tailles de fratries, prévalence de chaînes de filiation homogènes, etc.), tandis que la méthode de « travail de terrain virtuel » n'arrive pas à différencier entre les effets du souvenir généalogique, et des choix matrimoniaux sur les propriétés structurelles observables (par exemple, l'incidence qu'auraient la résidence virilocale d'une part, et la polygamie d'autre part, sur la prédominance des liens agnatiques dans un corpus généalogique).

La méthode préconisée consiste donc à combiner deux opérations de simulation – d'abord du réseau généalogique réel, puis de sa reconstruction à partir d'un ensemble de sous-réseaux égocentrés – en faisant appel, aux *deux* niveaux, aux *mêmes* facteurs sociologiques en

tant que déterminants aussi bien du comportement matrimonial que de la mémoire généalogique collective. Il sera ainsi possible, par l'application d'une morphogenèse « inverse », de partir d'un réseau exhaustif pour créer un réseau dégradé, incomplet, tel que l'on pourrait l'obtenir par la collecte empirique.

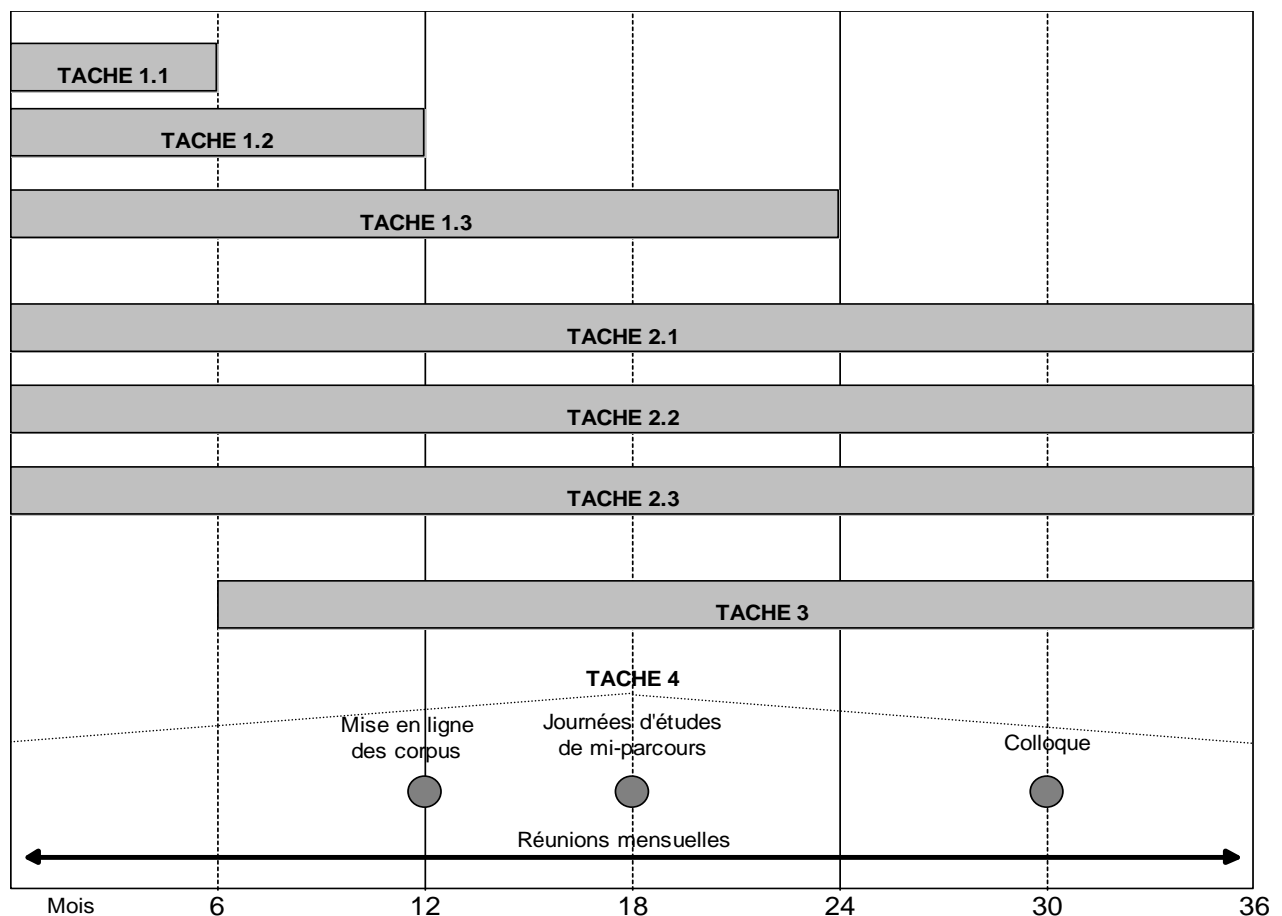
<u>Jalon J-3.1</u>	Algorithmes de simulation
<u>Livrable L-3.1</u>	Rapport technique sur la simulation de généalogies complètes
<u>Livrable L-3.2</u>	Rapport technique sur la simulation de généalogies incomplètes

### 3.3.4 TACHE 4 – IMPLEMENTATION, VALORISATION, USAGES & SERVICES

Voir la section 4 ci-dessous.

## 3.4. CALENDRIER DES TACHES, LIVRABLES ET JALONS

La figure 2 décrit l'agenda global du projet et l'agencement temporel des différentes tâches, tandis que le tableau 2 reprend la liste des livrables et jalons du projet.



**Figure 2 : Echancier des différentes tâches**

Tableau 2 : Tableau synthétique des livrables et jalons du projet

Tâche	Livrables et jalons	Date	Resp.
<b>0. Coordination et communication [coordination : INED]</b>			
	L-0.1. Site web du projet	M1	INED
	L-0.2. Rapports semestriels	M*	INED
<b>1. Collecte et codage des données [coordination : CEMAF]</b>			
	L-1.1. Corpus publiés codés	M6	CEMAF
	L-1.2. Corpus complétés codés	M12	CEMAF
	L-1.3. Analyse synoptique des corpus	M18	CEMAF
	L-1.4. Rapports des enquêtes de mémoire généalogique	M24	CEMAF
<b>2. Modélisation [coordination : INED]</b>			
	L-2.1. Rapport sur l'analyse comparative des structures matrimoniales dans des corpus empiriques	M*	CEMAF
	L-2.2. Rapport sur les outils conceptuels pour l'analyse des algèbre et topologie de circuits matrimoniaux	M*	CAMS
	L-2.3. Rapport sur les mesures et estimateurs statistique de la distribution des circuits matrimoniaux	M*	INED
<b>3. Simulation et validation [coordination : CAMS]</b>			
	J-3.1. Algorithmes de simulation	M30	CAMS
	L-3.1. Rapport technique sur la simulation de généalogies complètes	M36	CAMS
	L-3.2. Rapport technique sur la simulation de généalogies incomplètes	M36	CAMS
<b>4. Implémentation, valorisation, usage et services [coordination : INED]</b>			
	L-4.1. Mise en ligne des corpus codés	M12	CEMAF
	L-4.2. Logiciel ou module de simulation	M36	CAMS
	J-4.1. Ateliers de formation	M*	INED
	J-4.2. Journées d'études de mi-parcours	M18	INED
	J-4.3. Colloque final	M30	INED
	L-4.3. Publications	M*	

#### 4. STRATEGIE DE VALORISATION DES RESULTATS ET MODE DE PROTECTION ET D'EXPLOITATION DES RESULTATS

Les modèles et algorithmes développés seront implémentés, ou bien dans des **logiciels libres** indépendants, ou dans un **module de simulation** intégré au logiciel Puck (logiciel libre diffusé sous licence CeCILL, voir [www.kintip.net](http://www.kintip.net)). Ces logiciels seront rendus publiquement accessibles en ligne.

L'utilisation de ces techniques sera communiquée au public à travers un **manuel en ligne**, de même qu'une série d'**ateliers de formation** ouverts aux chercheurs et aux étudiants.

Un nombre important de corpus étudiés et leurs analyses statistiques seront rendus disponibles au public, sous forme anonymisée et protégée par une licence du type ScientificCommons, à travers un portail tel le site *KinSources* (projet commun de l'Université de Kent avec le groupe de recherche TIP).

Les résultats intermédiaires du projet seront présentés régulièrement dans le cadre de **colloques internationaux** tel que la conférence annuelle de *l'International Network of Social Network Analysis* (« Sunbelt Conference »).

Le projet assurera à mi-parcours des **journées d'études** portant sur les avancées méthodologiques et problèmes rencontrés. Ces journées d'études rassembleront des chercheurs internationaux compétents en modélisation, simulation et statistique.

Les résultats finaux du projet feront objet d'un **colloque international** organisé à Paris, réunissant chercheurs en études de parenté et en analyse de réseaux sociaux.

Les activités de recherche menées par les membres des différents partenaires institutionnels feront bien entendu l'objet de **publications**, individuelles et collectives, dans des revues et ouvrages spécialisés.

<u>Livrable</u> L-4.1.	Mise en ligne des corpus codés
<u>Livrable</u> L-4.2.	Logiciel ou module de simulation
<u>Jalon</u> J-4.1.	Ateliers de formation
<u>Jalon</u> J-4.2.	Journées d'études de mi-parcours
<u>Jalon</u> J-4.3.	Colloque final
<u>Livrable</u> J-4.3.	Publications

## 5. ORGANISATION DU PARTENARIAT

### 5.1. DESCRIPTION, ADEQUATION ET COMPLEMENTARITE DES PARTENAIRES

#### 5.1.1 PARTENAIRE 1 : INSTITUT NATIONAL D'ETUDES DEMOGRAPHIQUES (CHEF D'EQUIPE : FRANÇOIS HERAN)

L'INED s'intéresse de longue date à la collecte et à l'analyse des réseaux de parenté. Il a constitué des fichiers généalogiques très complets sur des groupes touaregs (Chaventré 1983, Cazes et Guignard 1991) et dogons. Depuis plus de vingt ans, l'INED entretient en Afrique de l'Ouest des « observatoires de population » qui recensent tous les événements démographiques dans des aires rurales où s'amorce la transition démographique (passage à une mortalité et à une fécondité réduites). Les fichiers individuels constitués à cet effet incluent des relevés généalogiques qui pourraient être complétés dans le cadre de ce projet. Par ailleurs, l'INED a participé à l'élaboration et à l'exploitation de la grande enquête de démographie historique dite des « 3000 familles » (connue également sous le nom d'enquête TRA), lancée par le laboratoire de démographie historique de l'EHESS avec l'appui du CNRS. L'expérience acquise, sur ce fichier généalogique construit à partir d'actes de mariage afin notamment de compléter les liens de parenté, a permis le lancement d'un « Pôle d'analyse des données relationnelles », porté conjointement par des statisticiens de l'INED et par les

ingénieurs du Laboratoire de démographie historique (LDH) de l'EHESS. Au cours de ses trois mandats successifs de directeur de l'INED, François Héran a fortement engagé les chercheurs de l'INED dans la réponse aux appels d'offres des programmes-cadres européens et des agences nationales de moyens (ANR, ANRS, AIRD). L'administration de l'INED est désormais rompue à la coordination de tels projets.

L'équipe est dirigée par le coordinateur du projet et directeur de l'INED, **François Héran** (voir section 5.2. ci-dessous).

**Arnaud Bringé** est responsable du service Méthodes Statistiques à l'INED depuis 2 ans. Il est spécialisé dans les méthodes d'analyse biographique et multiniveau. En parallèle, il s'intéresse depuis plusieurs années aux méthodes d'analyse des réseaux sociaux, et développe une activité sur cette thématique en collaboration avec Pascal Cristofoli, du Laboratoire de Démographie Historique de l'EHESS.

**Véronique Hertrich** est chargée de recherche à l'INED et responsable du projet « Dynamique démographique, conjugalité et parentalité en Afrique. Observation suivie d'une population rurale du Mali ». Ce projet s'articule sur une enquête biographique renouvelée tous les 5 ans, et sur un module de recueil généalogique de données par tradition orale.

**Pascal Cristofoli** est ingénieur d'études à l'EHESS, ingénieur associé à l'INED au sein du Service Méthodes Statistiques, où il effectue une veille technologique sur l'analyse de réseau, et développe en collaboration avec Arnaud Bringé des actions de valorisation sur cette thématique (Atelier de formation, organisation de journées scientifiques). Il est notamment l'animateur de l'Atelier « Analyse des données relationnelles », structure méthodologique se proposant de mettre en œuvre et promouvoir des études sur les réseaux sociaux et étudier les méthodologies utilisées dans ce domaine de recherche.

#### 5.1.2 PARTENAIRE 2 : CENTRE D'ANALYSE ET DE MATHÉMATIQUE SOCIALE (CHEF D'ÉQUIPE : CAMILLE ROTH)

Le CAMS est une unité mixte CNRS/EHESS rassemblant principalement mathématiciens, informaticiens et, dans une moindre mesure, sociologues autour de thématiques de modélisation et de formalisation des systèmes sociaux, au sens large. Il participe à la recherche mathématique internationale dans les domaines de compétence où il a acquis une notoriété à savoir la combinatoire et son algorithmique, la théorie des ordres partiels, les graphes, les représentations spatiales, la logique des structures du langage et les systèmes dynamiques. Dans le contexte des sciences humaines et sociales, le Centre développe de nouveaux modèles (combinatoires, probabilistes, dynamiques, informatiques) pour révéler les structures cachées, et a pu ces dernières années rassembler une équipe forte de chercheurs autour de l'étude des systèmes complexes. Le CAMS, enfin, est un lieu de débats épistémologique et historique sur la praxéologie mathématique.

**Camille Roth** est chargé de recherche CNRS en informatique après une brève expérience de

maître de conférences en sociologie, chercheur permanent au CAMS et associé au CREA et au CRESS (Centre of Research in Social Simulation, U. Surrey, GB). Ingénieur des Ponts & Chaussées (2002), il est titulaire d'un doctorat en sciences sociales de l'Ecole Polytechnique (2005) et d'un DEA de sciences cognitives (EHESS, 2002). Il a principalement effectué des travaux de modélisation mathématique de systèmes socio-sémantiques, en s'intéressant d'abord aux phénomènes de morphogenèse puis en se focalisant sur les phénomènes de diffusion de connaissances. Il a travaillé concrètement sur les réseaux de collaboration scientifique et sur les communautés en ligne (blogs, wikis, ...), en mettant en oeuvre des méthodes visant à reconstruire la structure taxonomique de communautés épistémiques ou à modéliser l'émergence de structures ou la dynamique des processus sous-jacents à partir de processus itératifs discrets. Ses recherches les plus récentes visent à appréhender les liens entre structure, motifs et phénomènes de viabilité, au sens large, de divers systèmes sociaux, tels que les réseaux d'entreprises ou les communautés virtuelles.

### 5.1.3 PARTENAIRE 3 : CENTRE D'ETUDES DES MONDES AFRICAINES (CHEF D'EQUIPE : MICHAEL HOUSEMAN)

Le CEMAF (UMR 8171 CNRS/Université Paris 1 – EPHE – Université de Provence) est un laboratoire interdisciplinaire (Histoire, Ethnologie, Anthropologie juridique, Sciences politiques) sous l'égide des directions scientifiques des sections 33, 38 et 40 du CNRS. Les recherches de ses membres portent sur l'ensemble du continent africain et s'ordonnent autour de quatre axes : (1) épistémologie et savoirs, systèmes de pensée et modes de représentation, (2) figures des pouvoirs et constructions du politique (VIII<sup>e</sup>-XXI<sup>e</sup> s.), (3) rites, cultes et métamorphoses du religieux, (4) patrimoines, environnement et cultures matérielles. Le CEMAF regroupe trois unités fondées sur des problématiques communes et sur un terrain partagé, et s'appuyant sur un réseau d'échanges déjà ancien intra- et interdisciplinaires. Un de ses pôles, le « Site Ivry » (anciennement le Laboratoire « Systèmes de pensée en Afrique noire »), regroupe des ethnologues dont plusieurs poursuivent des études sur les relations de parenté.

**Michael Houseman** est Directeur d'études à l'Ecole Pratique des Hautes Etudes (EPHE). Il a dirigé pendant huit ans le Laboratoire Systèmes de pensée en Afrique noire, devenu un des pôles du CEMAF. Il a publié plusieurs ouvrages et de nombreux articles sur la parenté et l'organisation sociale, ainsi que sur la modélisation des réseaux matrimoniaux, dont un chapitre sur l'analyse des réseaux matrimoniaux dans le *Sage Handbook of Social Network Analysis* (avec K. Hamberger et D. R. White, sous presse). Responsable du projet ANR (2005) « Traitement informatique des phénomènes de parenté » (TIPP), il a fondé l'équipe interdisciplinaire « Traitement Informatique de la Parenté » (TIP) qui rassemble des chercheurs en sciences humaines et sociales – ethnologues, historiens de la famille, sociologues – qui partagent un intérêt commun pour l'analyse systématique des relations de consanguinité et d'alliance.

**Klaus Hamberger** est économiste et anthropologue. Il est le concepteur du logiciel Puck (*Program for the Use and Computation of Kinship data*) et a publié plusieurs articles sur la modélisation et l'analyse formelle des réseaux de parenté. Il a notamment rédigé, avec Michael Houseman et Douglas R. White, l'article « Kinship Network Analysis » pour le *Sage Handbook of Social Network Analysis* (à paraître). Il a également effectué une recherche de terrain de 18 mois au Togo du Sud-Est et collecté un corpus généalogique important (disponible en ligne sur *Kinsource.net*). Il est actuellement chercheur contractuel au CEMAF dans le cadre du projet « Traitement Informatique des Phénomènes de Parenté » et est responsable d'un séminaire sur « Parenté et Espace » à l'EHESS.

**Isabelle Daillant**, chargée de recherche au CNRS, est ethnologue au Centre EREA (Enseignement et recherche en ethnologie amérindienne, Villejuif) du LESC (Laboratoire d'ethnologie et de sociologie comparative, UMR 7186, Nanterre). Spécialiste de parenté amazonienne, elle a effectué des enquêtes sur plus de 4 ans en Bolivie, chez les Chimane, auxquels elle a consacré un ouvrage et de nombreux articles. Elle est membre du groupe TIP (« Traitement informatique de la parenté ») et participe depuis 2002 à des groupes de recherche sur ces questions, dont, ces dernières années, le programme TIPP (« Traitement informatique des phénomènes de parenté »), dirigé par M. Houseman et C. Grange. Elle a contribué (en collaboration avec K. Hamberger) aux publications issues de cette recherche qui, associant ethnologues et historiens, a clairement fait apparaître la nécessité d'un travail approfondi sur le thème proposé ici.

**Laurent Gabail** est doctorant de l'université Paris Ouest Nanterre La Défense et du laboratoire CEMAF. Actuellement boursier du musée du quai Branly, il prépare une thèse sur l'organisation sociale des Bassari de Guinée. Fondé sur les données empiriques d'une recherche de terrain de quinze mois, son travail vise à mieux comprendre les rapports entre le système d'âge et les pratiques matrimoniales d'une société matrilineaire et virilocale.

**Michael Gasperoni**, doctorant, est allocataire de recherche de l'Université de San Marino et de l'EHESS, et moniteur à l'Université de Saint-Quentin-en-Yvelines, où il dispense un enseignement intitulé « Histoire et sciences sociales ». Il prépare actuellement une thèse portant sur l'Italie centrale, où il étudie, de manière comparative, la société majoritaire, chrétienne, et celle minoritaire, juive, saisies sous l'angle particulier de la parenté et des pratiques matrimoniales. Il a établi un corpus important et quasiment complet sur la population de San Marino (Gasperoni, 2008).

**Cyril Grange** est directeur de recherche au CNRS (Centre Roland Mousnier, UMR 8596). Ses travaux portent essentiellement sur l'histoire sociale des élites dans la France des XIX<sup>e</sup> et XX<sup>e</sup> siècles. Son HDR sur la grande bourgeoisie juive à Paris sous la III<sup>e</sup> République aborde particulièrement les questions de stratégies d'alliance, intra et extra-communautaires. Co-responsable, pour le Centre Roland Mousnier, du projet ANR « Traitement informatique des phénomènes de parenté » (TIPP), 2005-2009, il a coordonné le numéro des *Annales de Démographie Historique* consacré à « Histoire de la Famille et Analyse de Réseaux » (1/2005)

ainsi que le numéro « Les réseaux d'alliances, refonder l'analyse » avec Michael Houseman, à paraître en 2009.

**Olivier Kyburz** est ethnologue, maître de conférences à l'Université Paris-Ouest La Défense Nanterre. Il a dirigé le Département d'ethnologie, de préhistoire et d'ethnomusicologie durant quatre ans. Ses recherches de terrain portent sur la stratification sociale dans la moyenne vallée du Sénégal (Haalpulaar'en). Il participe à l'édition du Zuhur-al Basatin de Cheikh Moussa Kamara, une micro-histoire de la vallée du Ségégal (éditeur du second volume, à paraître). Actuellement en délégation au CNRS, il se consacre à la collecte et à l'analyse de matériaux généalogiques relatifs à plusieurs groupes endogames de cette société, Peuls, Aristocratie maraboutique (tooroBBe), pêcheurs... Ayant participé à l'ANR TIPP, il a publié avec Laurent Gabail « Hurons chez les Touareg une analyse comparée de deux réseaux matrimoniaux ».

## **5.2. QUALIFICATION DU COORDINATEUR DU PROJET**

**François Héran** est directeur de recherches de 1<sup>re</sup> classe, titulaire d'un doctorat de 3<sup>e</sup> cycle et d'un doctorat d'État en anthropologie de la parenté. Il dirige l'Institut national d'études démographiques depuis 1999. Il a été sociologue en Espagne et en Bolivie, avant de rejoindre l'Institut national d'études démographiques (INED) et l'Institut national de la statistique et des études économiques (INSEE), dont il dirige la division des enquêtes démographiques de 1993 à 1998. Il réalise des enquêtes nationales sur la sociabilité des Français (1982), la formation des couples (1984), les efforts éducatifs des familles (1992), la participation électorale (1995-1998), ainsi que l'enquête Famille associée au recensement de 1999. Il est l'auteur de nombreux ouvrages et articles scientifiques dont plusieurs sur l'analyse structurale et la modélisation des relations de parenté ; son dernier ouvrage, *Figures de la parenté*, est sous presse aux Presses universitaires de France. Il est depuis 2008 Président de la *European association for population studies* (EAPS). François Héran réunit ainsi des compétences relevant des trois champs de recherche sur lesquels s'appuie ce projet : la statistique démographique, la modélisation formelle et l'anthropologie de la parenté.

## **6. JUSTIFICATION SCIENTIFIQUE DES MOYENS DEMANDES**

### **6.1. PARTENAIRE 1 : INED**

- *Personnel*

Un/e post-doctorant/e sera recruté/e à mi-temps à partir de la deuxième année du projet afin de collaborer à la tâche 3, avec un profil « analyse de réseaux/statistiques appliquées » et une expérience significative de collaboration en SHS. Il/elle devra essentiellement étudier les conséquences de l'utilisation d'un réseau de parenté sur le calcul des différentes grandeurs de description d'un réseau, et sur l'élaboration d'un cadre conceptuel adapté pour pouvoir calculer la probabilité de formation d'un circuit matrimonial donné au sein des circuits de

parenté de même type. Il s'agit de 12 personnes.mois, soit environ 51600 euros (coût INED chargé).

Un administrateur/une administratrice du projet, chargé/e de la gestion administrative et financière, de la préparation des réunions, et du colloque final. Il s'agit de 6 personnes.mois pour un poste de type assistant de recherche, soit environ 21000 euros (coût INED chargé).

Des vacataires seront embauché/es pour le codage et la saisie des corpus. Il s'agit de l'équivalent de 3 personnes.mois, soit environ 8400 euros.

- *Missions*

Un budget sera prévu afin d'assister à des conférences en France et à l'étranger pour présenter les travaux issus du projet, d'un montant de 2000 euros pour l'équipe INED par an, soit 6000 euros au total.

La collecte des données (tâche 1) exigera des frais de mission pour un terrain extra-européen (9000 euros).

- *Autres dépenses de fonctionnement*

L'équipe aura besoin de matériel informatique afin de pouvoir intégrer le cadre conceptuel statistique dans l'application et tester sur les corpus du projets : il s'agira d'un ordinateur pour le post-doctorant recruté au sein du projet (~ 1500 euros) et un ordinateur utilisé par les différents membres de l'équipe dans le cadre de déplacements pour des réunions entre les équipes ou pour présentations extérieures, (~ 2000 euros), soit un total de 3500 euros.

L'institution coordinatrice aura en outre besoin de moyens pour l'organisation des journées d'études à mi-parcours (5000 euros) et du colloque final (10000 euros).

## **6.2. PARTENAIRE 2 : CAMS**

- *Personnel*

Un/e post-doctorant/e sera recruté/e pour la seconde moitié du projet afin de collaborer à la tâche 3, avec un profil « informatique/mathématiques appliquées » et une expérience significative de collaboration en SHS. Il/elle devra essentiellement se pencher sur la construction de réseaux de parenté réalistes exhaustifs (i) à partir d'hypothèses « micro-fondées », i.e. relatives au comportement individuel, et (ii) que ce soit en s'appuyant sur un corpus empirique partiel (morphogenèse partielle) ou bien en partant d'un réseau vide (morphogenèse totale). Il s'agit de 15 personnes.mois, soit environ 69000 euros (coût CNRS chargé).

- *Missions*

Un budget sera prévu afin d'assister à des conférences en France et à l'étranger pour présenter les travaux issus du projet, d'un montant d'environ 3000 euros pour toute l'équipe par an, soit 8000 euros au total.

- *Autres dépenses de fonctionnement*

L'équipe aura besoin de matériel informatique afin de mener à bien les simulations : il s'agira d'un ordinateur pour le post-doctorant recruté au sein du projet (~ 1500 euros) et un serveur de calcul léger dédié (~ 3500 euros), soit un total de 5000 euros.

### 6.3. PARTENAIRE 3 : CEMAF

- *Personnel*

Klaus Hamberger, concepteur du logiciel Puck et spécialiste de la théorie des réseaux de parenté, sera recruté sur un CDD à mi-temps sur la durée du projet. Il contribuera à la réflexion théorique (algèbre de circuits matrimoniaux), au développement des outils informatiques (modules de simulation intégré dans le logiciel Puck) et à la collecte de données (complétion d'un corpus généalogique important du Togo de Sud-Est). Il s'agit de l'équivalent de 18 personnes.mois, soit environ 82800 euros (coût CNRS chargé).

Des vacataires seront embauchés pour le codage et la saisie des corpus. Il s'agit de l'équivalent de 3 personnes.mois, soit environ 8400 euros (coût CNRS chargé).

- *Missions*

La collecte des données (tâche 1) exigera des frais de mission pour deux terrains européens (10000 euros au total) et quatre terrains extra-européens (30000 euros au total).

Un budget sera prévu afin d'assister à des conférences en France et à l'étranger pour présenter les travaux issus du projet, d'un montant de 2000 euros pour toute l'équipe par an, soit 6000 euros au total.

- *Autres dépenses de fonctionnement*

L'équipe aura besoin de matériel informatique pour la collecte et le codage des données (ordinateurs portables, appareils photos numériques, disques durs externes, etc.) pour un montant total de 6000 euros.

## 7. ANNEXES

### 7.1. REFERENCES BIBLIOGRAPHIQUES

Inclure la liste des références bibliographiques utilisées dans la partie « Etat de l'art » et les références bibliographiques des partenaires ayant trait au projet.

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.
- Bollobas, B. (1985), *Modern Graph Theory*, New York, NY. : Springer.
- Cazes, M.-H. et Guignard, E. (1991), « Les cercles d’alliance chez les Dogon islamisés de Tabi », in Héritier-Augé, F., Copet-Rougier, E. (dir.), *Les complexités de l’alliance, II : Les systèmes complexes d’alliance matrimoniale*, Paris, Éditions des Archives contemporaines.
- Chaventré, A. (1983), *Evolution anthropo-biologique d’une population touarègue. Les Kel Kummer et leurs apparentés*, Paris, PUF, INED
- Coult, A. D. et Randolph, R. R. (1965), « Computer Methods for Analyzing Genealogical Space », *American Anthropologist* 67 : 21-9.
- Daillant, I., et Hamberger, K. (sous presse), « Géométries dravidiennes. Une analyse comparative des structures matrimoniales de trois sociétés amazoniennes », *Annales de Démographie Historique*, 2008 (2), Dossier spécial « Les réseaux d’alliances. Refonder l’analyse ».
- Delille, G. (1985), *Famille et propriété dans le Royaume de Naples (XVe-XIXe siècles)*, Rome-Paris, Ecole Française de Rome.
- Dodds, P. S., Muhamad, R., Watts, D. J. (2003), « An Experimental Study of Search in Global Social Networks », *Science* 301 : 827-829.
- Dorogovtsev, S. N., Mendes, J. F. F., Samukhin, A. N. (2001), « Generic Scale of the "Scale-free" Growing Network », *Physical Review E* 63 (062101) : 1-4.
- Federico de la Rua, A. de (2004), « L’analyse longitudinale de réseaux sociaux totaux avec Siena », *Bulletin de méthodologie sociologique* 84, <http://bms.revues.org/index59.html>.
- Gabail, L. et Kyburz, O. (sous presse), « Hurons chez les Touaregs : une analyse comparée de deux réseaux matrimoniaux », *Annales de Démographie Historique*, 2008 (2), Dossier spécial « Les réseaux d’alliances. Refonder l’analyse ».
- Gasperoni, M. (2006), *Les pratiques matrimoniales juives dans l’Italie moderne (1450-1800). Perspectives de recherche*, Mémoire de DEA, Paris, EHESS.
- Gasperoni, M. (2008), *Popolazione, famiglie, parentela a San Marino in età moderna*, Quaderni del Centro Sammarinese di Studi Storici, Università di San Marino.
- Gilbert, J. P. et Hammel, E. A. (1966), « Computer Simulation and Analysis of Problems in Kinship and Social Structure », *American Anthropologist* N.S. 68 (1) : 71-93.
- Hamberger, K. (2009), *La Parenté Vodou. Organisation sociale et logique symbolique en pays ouatchi*, Thèse de doctorat, Paris, EHESS.
- Hamberger, K., et Daillant, I. (sous presse), « L’analyse de réseaux de parenté. Concepts et outils », *Annales de Démographie Historique*, 2008 (2), Dossier spécial « Les réseaux d’alliances. Refonder l’analyse ».

- Hamberger, K., Houseman, M., et White, D. R. (sous presse), « Kinship Network Analysis », in Carrington, P., et Scott, J. (dir.), *The Sage Handbook of Social Network Analysis*, London, Sage Publications.
- Hammel, E. (1976), *The SOCSIM demographic-sociological microsimulation program: operating manual*, Berkeley, Institute of International Studies, University of California.
- Héran, F. (1998), « De Granet à Lévi-Strauss », *Social Anthropology* 6 (1): 1-60 ; 6 (2) : 169-201 ; 6 (3) : 309-330.
- Héran, F. (sous presse), *Figures de la parenté*, Paris, PUF (trad. angl. en cours par H. Coleman et D. Mandelbaum, Oxford, Bardwell Press).
- Héritier, F. (1974), « Systèmes Omaha de parenté et alliance. Etude sur ordinateur du fonctionnement réel d'une société africaine », in Ballonoff, P.A. (dir.), *Genealogical Mathematics*, Paris-La Haye, Mouton.
- Holland, P. W. et Leinhardt, S. (1981), « An exponential family of Probability Distributions for Directed Graphs », *Journal of the Statistical American Association* 76 : 33-65.
- 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 Schweizer, T. et White, D. R. (dir.), *Kinship, Networks and Exchange*, Cambridge, Cambridge University Press.
- Kossinets, G. (2006) « Effects of missing data in social networks », *Social Networks* 28 : 247-268.
- Lang, H. (1995), *Demographic Micro-Simulation*. Paper presented at the Nanterre conference of the Working Group on Kinship and Computing, September, 1995.
- Laumann, E. O., Marsden, P. V., Prensky, D. (1989), « The boundary specification problem in network analysis », in Freeman, L. C., White, D. R., Romney, A. K. (dir.), *Research Methods in Social Network Analysis*, Fairfax, VA, George Mason University Press : 61-87.
- Leach, Edmund ([1952] 1968), « Les implications structurales du mariage matrilatéral des cousins croisés », in *Critique de l'anthropologie*, Paris, PUF.
- Little, R. J. A. et Schenker, N. (1995), « Missing data », in Arminger, G., Clogg, C. C., Sobel, M. E. (dir.), *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, New York, Plenum Press.
- Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenthat, I., Sheffer, M., Alon, U. (2004), « Superfamilies of evolved and designed networks », *Science* 303 (5663) : 1538-1542.
- Newman, M. E. J., Strogatz, S., Watts, D. (2002), « Random graphs models of social networks », *PNAS* 99 : 2566-2572.

- Nooy, de W., Mrvar, A., Batagelj, V. (2005), *Exploratory Social Network Analysis with Pajek*, New York, Cambridge University Press.
- Read, D. W. (1998), « Kinship based demographic simulation of societal processes », *Journal of Artificial Societies and Social Simulation* 1 (1), <http://www.soc.surrey.ac.uk/JASSS/1/1/1.html>.
- Rivière, P. (1969), *Marriage among the Trio : a principle of social organisation*, Oxford, Clarendon Press.
- Robins, G., Pattison, P., Woolcock, J. (2004), « Missing data in networks: exponential random graph (p\*) models for networks with non-respondents », *Social Networks* 26 : 257-283.
- Segalen, M. et Richard, P. (1986), « Marrying kinsmen in the pays bigouden sud », *Journal of Family History* 11 : 109-130.
- Selz, M. (1994), « Traitement informatique de données généalogiques : le logiciel GENPAR », *L'Homme* 130 : 129-136.
- Snijders, T. A. B. (2001), « The Statistical Evaluation of Social Networks Dynamics », *Sociological Methodology*, 31 : 361-395.
- Snijders, T. A. B. (2005), « Models for Longitudinal Network Data », in Carrington, P., Scott, J., Wasserman, S. (dir.), *Models and methods in social network analysis*, New York : Cambridge University Press.
- Stork, D. et Richards, W. D. (1992), « Nonrespondents in Communication Network Studies : Problems and Possibilities », *Group and Organization Management* 17(2) : 193-209.
- Tjon Sie Fat, F. (1990), *Representing Kinship: Simple Models of Elementary Structures*, Thèse de doctorat, Rijksuniversiteit Leiden.
- Travers, J. et Milgram, S. (1969), « An Experimental Study of the Small World Problem » *Sociometry* 32 : 425.
- Wasserman S. et Patterson, P. (1996), « Logit models and logistic regressions for social networks : I- An introduction to Markov graphs and p\* », *Psychometrika* 61 : 401-425.
- 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 Jorion, P. (1992), « Representing and Computing Kinship : A New Approach », *Current Anthropology* 33 (4) : 454-463.
- White, D. R. et Johansen, U. (2004), *Network Analysis and Ethnographic Problems: Process Models of a Turkish Nomad Clan*, Boston, Lexington Press.