

<TITRE> Explorer les réseaux à l'échelle de la triade : L'apport des modèles statistiques ERGM¹

Julien BRAILLY²
Fabien ÉLOIRE³
Guillaume FAVRE⁴
Alvaro PIÑA-STRANGER⁵

Résumé :

La statistique des réseaux sociaux, contrairement à la statistique classique, fait face à l'épineux problème de la dépendance des observations. Pour le résoudre, elle propose un nouveau type de modélisation statistique, les *Exponential Random Graph Models* (ERGM). À la base de ces modèles, se trouve la notion élémentaire de triade, discutée de longue date dans la littérature sociologique, en premier lieu chez Simmel. Cet article, qui vise, pour la première fois, à présenter en français, les fondements des ERGM, fait le point sur l'intérêt de l'approche triadique pour l'analyse des processus sociaux. Il donne un exemple détaillé de mise en œuvre des ERGM sur des données de réseau (celles d'un cabinet d'avocats d'affaires), et présente quelques travaux récents qui les mobilisent pour appréhender par exemple la sociabilité des adolescents ou les coopérations inter-organisationnelles.

Mots clefs : Réseaux Sociaux, Triades, ERGM, Méthodes statistiques, Configurations Structurales, Processus sociaux

Title: Exploring social networks at the triadic level: The contribution of ERGM statistical models.

Abstract:

In contrast to conventional statistical analysis, statistical models for social networks must account for dependencies between observations. To address this issue, a specific class of models has been developed, *Exponential Random Graph Models* (ERGM). The basis for these models is the fundamental idea of triads, a longstanding concept within sociology, as seen in Simmel's work. This article aims to present for the first time in French ERGM's theoretical foundations, and reviews the usefulness of the triadic approach for exploring social processes. Using a case study of a business law firm, the ERGM estimation process is presented in detail, followed by a review of recent research using ERGM.

Keywords: Social Networks, Triads, ERGM, Statistical methods, Structural configurations, Social processes

¹ Nous tenons à remercier Emmanuel Lazega pour ses conseils et encouragements, ainsi que pour nous avoir autorisés à utiliser ses données sur les avocats d'affaires. Nous tenons aussi à remercier Claire Lemercier, Dean Lusher, Gianluca Manzo, Mohamed Oubenal, Scott Viallet-Thévenin et Peng Wang pour leur relecture et précieux conseils.

² CTI-Swinburne University of Technology, H23 PO Box 218, Hawthorn, Victoria 3122, Australia ; CSO-Sciences Po-Paris ; j.brailly@cso.cnrs.fr

³ Université de Lille1, CLERSE – UMR8019, Cité Scientifique, 59655 Villeneuve d'Ascq ; fabien.eloire@univ-lille1.fr

⁴ LISST-CERS, UMR 5193 CNRS, Université Toulouse Jean Jaurès, 5, allées Antonio-Machado, 31058 Toulouse cedex 9, France ; guillaume.favre@univ-tlse2.fr

⁵ CREM, Université de Rennes 1, France et ICSO, Universidad Diego Portales, Santiago, Chile ; alvaro.pina-stranger@univ-rennes1.fr

<TEXTE>

L'analyse des réseaux sociaux est une méthodologie qui oriente les sociologues dans une voie de recherche originale, où les structures sociales sont appréhendées non pas seulement à travers les attributs des individus (sexe, âge, PCS, niveau de diplôme, etc.) mais aussi à travers les réseaux d'interactions dans lesquels ils s'inscrivent à l'échelle méso-sociale. Ces méthodes s'inscrivent et trouvent leurs racines dans de nombreuses approches sociologiques depuis celle de Simmel pour qui la triade, c'est-à-dire l'ensemble des relations entre trois acteurs, est la structure élémentaire du social, génératrice de « formes » sociales.

Jusqu'au début des années 2000, en raison de l'absence de modèles statistiques spécifiques, l'exploration des réseaux à l'échelle triadique était avant tout métaphorique ou qualitative. L'apparition au début des années 2000 des *Exponential Random Graph Models* (ERGM) comble ce vide et ouvre de nouvelles perspectives. Si la littérature internationale compte nombre d'études utilisant ces modèles, ils restent peu connus et utilisés en France, et surtout rarement justifiés sociologiquement. Or la focale sur les triades, permise par cette classe de modèles, peut être replacée dans une tradition simmeliennne, et plus généralement dans celle de l'analyse des réseaux sociaux. Fort de ce constat, l'objectif de cet article est de présenter les ERGM et de montrer leur intérêt pour observer des régularités et révéler des processus sociaux au sein de collectifs.

L'arrivée tardive de ces modèles s'explique notamment par la spécificité des données de réseaux d'un point de vue statistique. Alors que la statistique classique part du postulat d'indépendance des observations, la statistique des réseaux sociaux étudie au contraire les interdépendances entre les individus à travers les relations qu'ils entretiennent entre eux. En manipulant des objets mathématiques tels que des matrices et des graphes, elle fait face à l'épineux problème de la dépendance des observations.

Afin de mieux prendre en compte cette dépendance, et donc de tenir compte du fait que « le lien crée du lien »⁶, les premiers modèles statistiques de réseaux basés sur la fonction exponentielle ont vu le jour à la fin des années 80 (appelés modèles p1 et p2) à la demande spécifique de sociologues des réseaux. À la suite de ces modèles, Stanley Wasserman et Philippa Pattison (1995) ont proposé les premiers ERGM, basés sur les travaux fondateurs d'Holland et Leinhardt (1976). Plutôt que de décomposer le réseau en un ensemble de dyades, la logique des modèles ERGM consiste alors à réinsérer celles-ci dans leur voisinage relationnel, à savoir les triades, afin de repérer lesquelles de ces sous-structures élémentaires reviennent le plus fréquemment, et sont donc les plus structurantes dans le réseau étudié. Dit autrement, l'enjeu de ces modèles est d'expliquer la structure du réseau, là où la plupart des autres modèles pour les données de réseaux se centrent sur les attributs des acteurs. Depuis lors, les ERGM ne cessent de connaître des améliorations, que ce soit au plan statistique ou au plan informatique, comme en témoigne la récente parution d'un ouvrage méthodologique sur le sujet codirigé par Lusher, Koskinen et Robins (2013).

Pour les sociologues, une des évolutions les plus significatives est certainement la possibilité de combiner, au sein de mêmes modélisations, les attributs sociaux, classiques en sociologie (du type sexe, âge, diplôme, opinions, etc.), et les données de réseaux, ce qui constituait auparavant une limite essentielle de ces approches. Ce faisant, ces modèles appartiennent à la grande famille des outils statistiques à même de « révéler » des structures « cachées ». Leur angle d'approche réticulaire offre au sociologue un point de vue renouvelé sur les phénomènes sociaux à l'échelle locale, par exemple en termes de réciprocité, cliques ou transitivité. Et ce, sans pour autant se détacher des problématiques, plus standard, en termes de classes sociales ou de structures sociales.

⁶ Par exemple, dans les situations où l'ami de mon ami devient mon ami aussi.

Dans une première partie, nous présenterons l’itinéraire de la notion de triade dans la littérature des sciences sociales ; dans une deuxième, nous décrirons les modélisations de type ERGM en tant que méthode statistique et la conception du social sur laquelle elles reposent ; dans une troisième, nous suivrons la réalisation et l’interprétation d’un modèle ERGM sur la base de données empiriques afin de montrer la plus-value d’une telle modélisation. La dernière partie présentera quelques contributions utilisant les modèles ERGM, révélant des mécanismes triadiques illustrant des processus sociaux plus ou moins complexes ainsi que les dernières avancées et limites de cette classe de modèles.

Encadré 1 : Terminologie et définitions

Dans la suite de cet article, nous allons utiliser de nombreux concepts propres à l’analyse des réseaux sociaux. Les principaux sont les suivants (pour plus de détails voir les ouvrages de Degenne et Forsé (1994 [2004]), Wasserman et Faust (1994) ou Lazega (1998 [2014])) :

- **Réseau** : d’un point de vue sociologique, un réseau social se compose d’*entités sociales* (individus, organisations, etc.) et de *relations* sociales entre ces acteurs telles que l’amitié ou l’échange de conseil. Ces relations peuvent être *orientées* (lorsque le sens des relations entre les entités sociales est pris en compte) et *multiplexes* (circulation de plusieurs relations entre ces entités).
- **Graphe** : traduit en langage mathématique, un réseau social correspond à un *graphe*, composé de *nœuds* et d’*arcs* ou d’*arêtes*. Il peut être décomposé en *sous-graphes* (ou *configurations structurales*) plus simples de n acteurs. On parlera alors de sous-structures d’ordre n (une sous-structure d’ordre 2 correspond à une dyade, d’ordre 3 à une triade, etc.).
- **Densité** : la densité d’un réseau correspond au nombre de liens observés (par l’enquête empirique) divisé par le nombre de liens possibles (si tout le monde se connaissait dans le réseau).
- **Degré** : Le degré d’un acteur correspond au nombre de lien qu’il possède. Pour des réseaux orientés, il est possible de différencier le *demi-degré intérieur* d’un individu ou *popularité* correspondant au nombre de liens reçus par l’individu, du *demi-degré extérieur* ou *activité*, correspondant au nombre de liens déclaré par l’individu.
- **Dyade** : une dyade se compose de deux individus, et de l’ensemble des liens qui existent (ou non) entre eux. Dans un réseau orienté, on parle de dyade *réciproque* lorsque A déclare être en relation avec B, et B déclare, lui aussi, être en relation avec A.
- **Triades** : une triade se compose de trois individus et de l’ensemble des liens qui existent (ou non) entre eux. De manière combinatoire, il existe quatre types de triades différents pour des réseaux non orientés (allant de la triade vide, sans lien entre les trois acteurs) à la triade pleine (ou tous les acteurs sont en relations). Pour les réseaux orientés, seize triades sont à différencier. Dans la suite de l’article, nous mentionnerons souvent trois triades particulières (pour des réseaux orientés) : le chemin transitif entre deux acteurs ou *chemin de pas-2* (A est en relation avec B, B est en relation avec C) ; ce chemin se ferme lorsque A et C sont aussi en relation. Suivant le sens de cette relation on parlera de *triade transitive* (A est en relation avec C) ou de *triade cyclique* (C est en relation avec A).
- **Homophilie** : tendance des individus ayant un attribut particulier en commun à être en relation (suivant l’adage « qui se ressemble s’assemble »).

<IT1>1. La triade en tant que sous-structure élémentaire pour l’analyse du social

Les modèles ERGM analysent les réseaux en prenant notamment comme unité d’analyse la triade (mais pas exclusivement). Ce faisant, ils s’inscrivent dans la tradition sociologique étudiant les réseaux à cette échelle, par exemple chez Simmel pour qui la triade constitue une

structure élémentaire pour l'émergence de formes sociales, ou chez d'autres auteurs pour qui elle constitue une métaphore des structures sociales.

<IT2> *La triade comme métaphore des structures sociales*

Simmel a souligné l'importance, dans une relation, de ce qu'il appelle le « *chiffre trois* » (1908 [1999], p.128). Pour lui, la société ne naît pas à l'échelle de deux mais de trois individus, car c'est à cette échelle qu'émerge l'entité supra-individuelle qu'est le groupe. Dans son raisonnement, si la dyade ne constitue pas un niveau supra-individuel, c'est parce que si deux personnes s'associent pour faire quelque chose ensemble, la disparition de l'une des deux fait disparaître leur association ; par contre, l'ensemble formé par les relations entre trois individus fait naître une nouvelle entité qui perdure suite à la disparition de l'un des membres. Dans ce second cas, le groupe dépasse l'individu et ce sont ces relations, basées sur la présence de trois individus, qui font apparaître ce que Simmel appelle des « formes sociales », telles que la concurrence, la domination, l'amour ou le conflit.

Ainsi, il présente la concurrence comme une « *lutte indirecte* » dont l'enjeu est extérieur aux concurrents car il n'appartient à aucun d'eux : chacun lutte pour ce même enjeu mais sans forcément chercher à se nuire directement. La concurrence ayant souvent pour objectif de séduire un tiers, la disparition d'un concurrent est bien souvent un effet collatéral de la lutte victorieuse de l'un d'eux. On retrouve là l'importance de la triade : c'est le fait de lutter pour l'amour d'une même personne, ou pour vendre un bien au même acheteur, qui fait naître la forme sociale *concurrence*. Les relations entre trois individus font alors apparaître un nouvel élément : celui du *rôle*.

Dans « Les formes de la socialisation », Simmel examine les différentes fonctions du tiers⁷. Parfois, le tiers est l'élément extérieur qui constitue le fondement de l'unification de la dyade : il illustre cela par l'exemple de l'enfant dont « *la fonction est de maintenir la cohésion* » au sein du mariage monogamique. Parfois, le tiers est un médiateur auquel deux parties font appel, ou se réfèrent par la force des choses, en cas de conflit manifeste autour de revendications contradictoires : Simmel discute la question des conditions de l'impartialité du médiateur, et de ses capacités d'arbitrage impliquant une confiance des deux parties. Parfois, le tiers est un *tertius gaudens*, susceptible de tirer avantage d'un conflit entre deux parties, tel, dans un régime parlementaire, un petit parti qui négocie son soutien à l'une des deux formations dominantes en opposition. Enfin, parfois, le tiers s'efforce de jouer l'une contre l'autre des forces au départ unies contre lui : c'est le *divide et impera* (diviser pour régner), illustré par le cas de la tribu vainqueur qui impose à la tribu vaincue de se partager en deux, et de se choisir deux chefs.

Ces trois cas de figure illustrent le saut qualitatif que représente le passage du chiffre deux au chiffre trois, saut qui est beaucoup moins prononcé lorsqu'on passe du trois au quatre. En effet, la triade est une forme sociale particulière, dont le sens ne s'épuise pas dans la description des trois dyades successives qui la composent. Elle apparaît plutôt comme une forme sociale élémentaire pour des groupes d'importance supérieure : alors que l'équilibre dans la vie d'un couple se trouve entièrement transformé par l'arrivée du premier enfant, le changement est beaucoup moins prononcé avec l'arrivée du second, notamment parce que la famille a atteint la forme supra-individuelle d'un groupe (Simmel, 1908 [1999], p.132 ; Caplow, 1968 [1971], p.37).

Chez Lévi-Strauss (1967) aussi, on trouve une différenciation entre les situations où deux partenaires interviennent exclusivement, qu'il nomme échange restreint, et des structures cycliques d'une plus grande complexité, mais triadiques dans leur forme la plus basique, qu'il

⁷ « *Il me semble que la triade en tant que telle engendre trois formes typiques de regroupement qui, (...) ne sont pas possibles avec deux éléments* » (p. 131).

nomme échange généralisé et qui autorisent une multiplicité de combinaisons possibles. Ici, l'échange est indirect entre au minimum trois individus, de sorte que : A donne à B qui donne à C qui rend à A, afin que le cycle se referme sur lui-même. Cependant, Lévi-Strauss suggère aussi qu'« *au cycle principal se joindront des cycles secondaires* ». Les triades élémentaires sont alors replacées dans un ensemble plus vaste, un réseau d'ordre supérieur. Et c'est alors tout un système d'échanges qui se découvre au fil de la reconstitution des cycles.

En anthropologie, une autre référence est la *Kula* mélanésienne, « *forme d'échange intertribal de grande envergure* », étudiée par Malinowski (1922 [1989], p.139). La carte V du circuit *Kula* (p.140), évoque parfaitement un tel système d'échanges où les traits en pointillé, qui relient entre elles des îles situées en Nouvelle-Guinée, représentent la circulation de deux sortes d'objets d'apparat précieux. Le système *Kula* constitue une véritable découverte ethnographique car les participants à cette « *institution ample, complexe et pourtant bien réglée* » n'ont, écrit Malinowski, « *aucune conscience de ses lignes directrices. (...) Aucun indigène ne se fait une idée claire de la Kula en tant que vaste institution sociale organisée (...) : il vit dans la Kula mais ne parvient pas à prendre le recul nécessaire pour la voir dans son ensemble* » (pp.141-142). L'analogie avec le réseau social est ici frappante.

Le recours à une structure tripartite afin d'illustrer des processus sociaux s'agrégeant et expliquant une partie des structures sociales est aussi mobilisée par Steiner (2009), qui réfléchit à la question de la transplantation d'organe dans les sociétés modernes, c'est la notion de « *sacrifice* », plutôt que celle de « *don* », qui lui semble rendre le mieux compte de ce phénomène. Pour lui, ce ne sont pas deux, comme dans le don archaïque interpersonnel chez Mauss (1950), mais bien trois participants qui interviennent dans le don impersonnel entre inconnus que les progrès chirurgicaux rendent possible aujourd'hui (le donneur, le receveur, le chirurgien).

Enfin, on trouve aussi une forme sociale tripartite chez Bourdieu (2012). Afin d'expliquer la division du travail de domination au sein de l'Etat dynastique, il mobilise l'image d'un triangle où, à chaque extrémité, se trouvent des personnages en rivalité pour le pouvoir dynastique 1) le roi, 2) les frères du roi, 3) les ministres. Cette structure triadique permet de comprendre pourquoi, dans beaucoup d'anciens empires, les bureaucrates ou hauts fonctionnaires sont des « *parias* », exclus de reproduction (politique, biologique, etc.), souvent voués au célibat : eunuques, prêtres, étrangers, renégats, esclaves propriété de l'Etat, ou mercenaires.

<IT2>De la métaphore des structures sociales à l'unité d'analyse

L'analyse à l'échelle triadique apparaît porteuse de sens sociologiquement en ce qu'elle permet de décrire et analyser élémentairement les phénomènes collectifs. Elle présente également l'intérêt de pouvoir être formalisée pour rechercher des régularités méso-sociales, à l'échelle du réseau. Ainsi, les analyses de réseaux sociaux ont contribué à la compréhension du rôle des triades du point de vue des relations sociales.

Un travail de synthèse spécifiquement consacré à « *la triade comme système social* » est réalisé, en 1968, par Caplow, lors duquel il fait le lien entre les réflexions pionnières de Simmel et les travaux des psychosociologues des années 1950. Dans la définition qu'il donne de la triade, il insiste sur une caractéristique essentielle qui est « *sa tendance à se diviser pour former une coalition de deux de ses éléments contre le troisième* » (p.10). Cette tendance s'inscrit dans un processus relevant « *d'une sorte de géométrie sociale* » où la structure de chaque coalition possible est liée à la « *force* » relative de ses éléments. À partir de là, Caplow bâtit une typologie en huit différentes formes de triades, fondée sur les forces de ses entités constitutives. L'ouvrage de Caplow passe ensuite longuement en revue l'ensemble des cas où la triade est génératrice de connaissance des systèmes sociaux, que ce soit en matière

d'expérimentation psychologique, d'observation du comportement animal, ou de compréhension des organisations, des structures familiales ou de la marche de l'Histoire.

Bien qu'il ne fasse pas référence à Caplow, l'article de Granovetter consacré à « *la force des liens faibles* » (1973) témoigne aussi du caractère fructueux des réflexions sur les triades. Il s'inspire de la théorie psychologique de la dissonance cognitive de Heider et Newcomb. Il met l'accent sur l'importance des triades comme formes élémentaires exerçant une pression sur les dyades qui les composent, de sorte que : quand les relations A-B et A-C existent et qu'elles sont des liens forts (des amis, par exemple), il est peu probable qu'une relation entre B et C ne finisse pas un jour par se former. Pour Granovetter, il existe donc une triade « *interdite* », c'est la triade A-B, A-C où A-B et A-C sont des liens forts et où B-C n'existe pas. Les liens faibles ne peuvent donc, pour Granovetter, qu'exister qu'entre deux groupes non connectés dans un réseau expliquant par-là pourquoi ces derniers sont plus efficaces en termes de recherche d'emploi.

Parmi les travaux importants qui prennent la triade pour unité d'analyse, il faut aussi évoquer ceux de Burt (1995), qui reprend d'ailleurs à son compte la notion de *tertius gaudens* de Simmel. Il considère que la « *triade interdite* », où A fait l'intermédiaire entre B et C qui ne se connaissent pas, est d'autant plus importante à repérer dans un réseau personnel qu'elle permet à A de bénéficier d'une position avantageuse, d'autonomie, du fait du « *trou structural* » existant entre ses contacts B et C. Cette position de « *pont* » entre contacts dits « *non redondants* » (car ils ne sont pas reliés) est analysée comme stratégique car conduisant à une meilleure efficacité et performance.

On voit bien comment les analyses triadiques ont inspiré les sociologues des réseaux sociaux. Les régularités triadiques observées dans un réseau témoignent alors de la présence de mécanismes relationnels locaux qui produisent le réseau observé. De cette manière l'analyse des triades permet, en quelque sorte, de saisir comment sont produites les « *formes sociales* » décrites par Simmel. Il ne s'agit donc pas de regarder chaque relation indépendamment, mais d'étudier le contexte relationnel de celles-ci (par exemple comment les relations s'insèrent dans des sous-structures plus complexes telles que les triades) : tel est justement l'enjeu des modèles ERGM.

<IT1>2. Les modélisations ERGM comme méthode statistique permettant de se centrer sur les triades

D'un point de vue statistique un réseau social constitue un objet à la fois pauvre et riche : pauvre, parce que le réseau est avant tout une construction du chercheur, dont la frontière est toujours floue et dépendante du processus social étudié (Laumann et *al.*, 1989 ; Éloire et *al.*, 2011), et qu'il serait illusoire de vouloir prétendre cerner, dans un même mouvement ou modèle, l'ensemble des tenants de la création, du maintien et de la dissolution des liens entre acteurs, alors même que chacun est le fruit d'une histoire propre. D'un autre côté, un réseau est riche statistiquement parlant, parce qu'il permet de reconstituer systématiquement l'ensemble des liens observés entre un ensemble d'acteurs, à un instant *t*, quelle que soit leur ancienneté ou le contexte de leur construction, afin d'y rechercher des processus sociaux récurrents. Ces processus témoignent en partie de ce que le réseau est autant structuré par les acteurs lorsqu'ils créent ou rompent des liens, que structurant pour eux lorsque le réseau oriente, influence, et détermine en partie leurs comportements.

<IT2>Quand le lien crée du lien : endogénéiser les mécanismes à l'échelle méso-sociale

La capacité à inscrire l'action individuelle dans le cadre d'une approche qui prend en compte le poids des structures sociales constitue le problème premier d'une modélisation de type non déterministe des données de réseaux. Ainsi, la systématisation de la collecte de données par un questionnaire sociométrique et des générateurs de noms (Éloire, Penalva-Icher, Lazega, 2011 ; Lazega 1998 [2014]) répond à l'idée de dépendance entre les différentes observations. Dans ce cadre, la stratégie d'analyse des ERGM consiste à contextualiser les dyades. Le postulat est qu'un réseau n'est pas qu'un ensemble de dyades juxtaposées et indépendantes, mais que sa morphologie globale résulte, au moins en partie, d'effets structuraux pouvant être saisis à l'échelle des triades et de leur voisinage relationnel. Ce dont il s'agit ici, c'est de la dimension dite *endogène* du réseau, caractérisée par le fait que le lien crée du lien, de sorte que, par exemple, l'existence d'un lien entre deux acteurs est liée à l'existence d'un effet structural de fermeture des triades par cycle. Cela implique de postuler que le réseau est généré par des processus stochastiques particuliers, à savoir des processus sociaux influant sur la création des liens, et donc à le considérer comme auto-organisé (Robins *et al.*, 2005 ; Robins *et al.*, 2007).

Ainsi, l'objectif de la modélisation statistique consiste à comprendre la forme globale du réseau observé, et plus particulièrement, les déterminants de la création des liens, en mobilisant des effets *endogènes* basés uniquement sur la structure de ce réseau. Les attributs sociaux des acteurs, dits *exogènes* au réseau, ne seront introduits que dans un second temps, une fois évacués les effets purement structuraux. En écho aux erreurs écologiques et atomistes de Robinson (1950) qui surestiment soit le contexte, soit les caractéristiques individuelles dans l'observation d'un phénomène social, cette endogénéisation doit permettre d'éviter une *erreur structurale*.

Par exemple, au sein d'un réseau d'acteurs de deux types, A et B : il ne suffit pas d'observer qu'il y a plus de liens entre les acteurs de type A qu'entre les acteurs de type B pour conclure à une homophilie plus forte pour les premiers. Ce résultat peut venir du fait qu'une partie des acteurs de type A ont un autre attribut A1 expliquant la majorité des liens. Cela souligne l'intérêt d'avoir un modèle permettant de mettre en compétition les différents attributs. Ce résultat peut aussi provenir du fait que les acteurs de type A ont beaucoup plus de liens que les acteurs de type B et que, en parallèle, le réseau observé soit caractérisé par un nombre important de triades d'un même type. Il est donc possible, en prenant en compte effets structuraux et attributs, d'obtenir le résultat inverse : il n'y a pas de tendance à l'homophilie entre les acteurs de type A mais il y en a une entre les acteurs de type B. Schématiquement, ici, pour tester l'homophilie, il faut corriger les liens observés de l'influence des attributs A1, de la spécificité de la distribution de degré (certains acteurs ont beaucoup de liens et d'autres non) et de la tendance à la fermeture triadique.

<IT2> *L'hypothèse de dépendance : des structures locales à la morphologie globale*

Il est possible de résumer comme suit l'idée théorique des ERGM (Robins et Lusher, 2013) : les réseaux sociaux comportent une dimension émergente et auto-organisée traduisant l'existence de plusieurs mécanismes structuraux et endogènes en action (le lien crée du lien) et une dimension déterminée et exogène (l'attribut explique le lien). Pour bien saisir la façon dont s'effectue l'endogénéisation de la structure au sein de ces modèles, il convient de revenir sur les quatre grandes hypothèses de dépendance structurale locale qui se sont succédées au cours de l'évolution historique et conceptuelle de ces modèles (Koskinen et Daraganova, 2013) : 1) l'hypothèse de Bernoulli, 2) de dépendance dyadique, 3) de dépendance markovienne, 4) de dépendance du modèle du circuit social.

La première hypothèse dite de Bernoulli est la plus simple et renvoie à la modélisation d'Erdős-Rényi (1959) selon laquelle la probabilité d'existence d'un lien correspond à la

densité du réseau et est la même pour tous les liens de ce réseau. Selon cette conception, toutes les dyades potentielles ont la même chance de se réaliser, de la même façon que l'on a une chance sur deux de gagner à pile ou face (avec une pièce non truquée). Cependant, on comprend logiquement que si, A déclare être ami avec B, il y a de fortes chances que B déclare être ami avec A, ce dont il convient de tenir compte dans la modélisation statistique.

La deuxième hypothèse dite de dépendance dyadique revient à intégrer cet effet fondamental de réciprocité : le lien entre A et B est dépendant du lien entre B et A. Il s'agit de l'un des effets les plus forts et les plus réguliers observé en analyse des réseaux sociaux. Mais comme nous l'avons vu précédemment avec la « triade interdite », il existe des effets structuraux triadiques poussant à la fermeture des triades. L'existence d'un chemin de longueur deux entre deux acteurs (c'est-à-dire, A est en relation avec B qui est en relation avec C) constitue un contexte relationnel spécifique qui favorise l'existence d'un lien entre ces acteurs. Dans ce cas, les liens sont dépendants entre eux dès lors qu'ils ont un acteur en commun. Si A est l'ami de B et B l'ami de C, on peut penser que A a de fortes chances d'être l'ami de C en raison du fait qu'ils soient reliés par B.

On en arrive alors à la troisième hypothèse dite de dépendance markovienne selon laquelle deux liens sont indépendants (conditionnellement au reste du réseau) s'ils n'ont aucun nœud en commun. Dans cette perspective, il convient d'étudier les configurations structurales d'ordre trois, les triades. Cette idée se base sur le postulat de Holland et Leinhardt (1976) selon lequel on peut comprendre les processus en jeu dans un réseau en s'intéressant uniquement aux structures locales, dans la mesure où ces dernières sont dépendantes entre elles⁸. Pour simplifier, on pourrait dire que les configurations locales d'un réseau (la distribution des triades) conduisent à des réseaux bien particuliers dont les caractéristiques globales peuvent être en partie saisies par l'agencement relatif des caractéristiques locales (Robins et al. 2005). Ce théorème statistique invite, d'une certaine manière, à faire le lien entre le niveau individuel (et la dyade) et le niveau structural (le réseau dans son ensemble), en étudiant uniquement le voisinage relationnel des acteurs (les configurations structurales telles que les triades). Par exemple, dans le cas d'un réseau formé par deux zones, un cœur très dense où les acteurs ont beaucoup de relations entre eux et une périphérie très marginale, on peut supposer que ce réseau sera caractérisé par un nombre important de triades (entre les acteurs du centre)⁹.

Bien qu'il existe des pressions sociales à la fermeture des triades, il n'est pas rare d'observer que deux individus ayant beaucoup d'amis en commun ne soient pas amis. L'absence de lien social entre deux personnes partageant un nombre important de relations n'est pas sans raison. La quatrième hypothèse, dite du circuit social, permet de prendre en compte cet effet. Elle propose de « contextualiser » les triades en prenant en compte la structure globale du réseau. Pattison et Robins (2002) proposent d'étendre le concept de dépendance markovienne à celui de dépendance conditionnelle partielle : deux liens sont indépendants, non seulement en fonction des nœuds qu'ils partagent, mais aussi en fonction de la structure des liens du reste du graphe (Snijders et al., 2006 ; Robins et al., 2007). La logique de raisonnement est toujours triadique, mais on intègre ici les effets de la multiplicité

⁸ Pour un nombre d'acteurs donné, la prise en compte des valeurs des différentes triades conduit à un sous-ensemble spécifique de réseaux possibles (Besag, 1974 ; Frank et Strauss, 1986). Or, comme le montre le théorème d'Hammersley-Clifford (1971), si le réseau vérifie l'hypothèse de dépendance précédente, les propriétés globales du réseau et celles du voisinage des points sont les mêmes. Il suffit donc de connaître uniquement le voisinage des points pour décrire la morphologie du réseau.

⁹ Pour un réseau non orienté de 10 acteurs avec uniquement 10 liens entre 5 acteurs (ils sont donc tous reliés directement entre eux), il suffit de compter le nombre de triades (30) pour comprendre qu'il s'agit d'un réseau avec un cœur parfaitement connecté et une périphérie complètement isolée. Le fort nombre de triades (conditionnellement au nombre de lien et d'acteurs) explique la morphologie du réseau : un cœur parfaitement relié et une périphérie isolée.

des chemins possibles comprenant, dans l'absolu, l'ensemble du réseau : si deux individus sont reliés par un nombre important de chemin de 2-pas (ils sont en relation avec les mêmes tiers), et n'ont pas de relation, cela a une signification. Cette dernière hypothèse de dépendance implique que l'on va modéliser le réseau à partir des structures suivantes : la densité, la distribution du degré des acteurs, les triades et les structures de logique triadique mais d'ordre supérieur.

Ces principales hypothèses de dépendance structurale sont représentées sur la figure suivante.

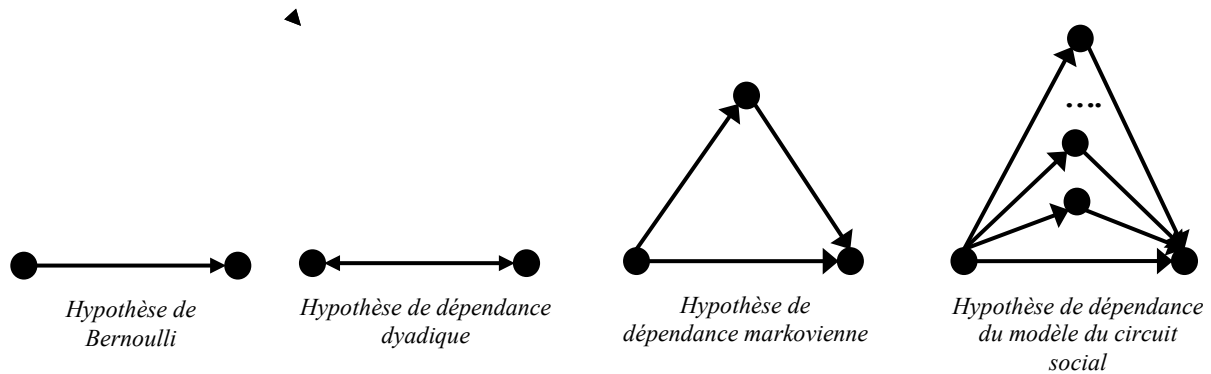


Figure 1 : Différentes hypothèses de dépendance structurale

La formule générale du modèle peut s'écrire sous la forme suivante (Robins et Lusher, 2013) :

$$P_{\theta}(Y = y) = \frac{1}{\kappa(\theta)} \cdot \exp\left(\sum_p \theta_p \cdot Z_p(y)\right)$$

Avec y le réseau observé par l'ensemble des réseaux possibles ayant le même nombre d'acteurs Y , $Z_p(y)$ la statistique de réseau de la structure locale de type p (dyade, triade, ...) et θ_p le paramètre associé. $\kappa(\theta)$ est la constante de normalisation correspondant à l'ensemble des graphes possibles avec le même nombre de nœuds selon les configurations structurales prises en compte dans le modèle¹⁰. Ainsi, ce qui est estimé est bien la probabilité d'observer le réseau réel parmi l'ensemble des réseaux possibles, et non la probabilité d'existence d'un lien.

<IT2>Le système social comme milieu : de l'hypothèse d'homogénéité aux effets exogènes

L'équation précédente implique qu'il y ait autant d'estimateurs que le nombre des configurations prises en compte, ce qui d'un point de vue computationnel est irréalisable pour des réseaux de taille de plus de 30 acteurs¹¹. Afin de simplifier le modèle, est alors posée l'hypothèse d'homogénéité des configurations isomorphiques selon laquelle l'impact des sous-structures est le même, dans l'émergence du réseau global, quels que soient les acteurs

¹⁰ En ce qui concerne la question de l'estimation statistique de tels modèles, deux choix s'avèrent possibles, soit les techniques d'estimation basées sur la pseudo-vraisemblance (équivalentes algébriquement à une régression logistique), soit celles basées sur les lois des familles exponentielles (sous le nom d'*Exponential Random Models*), qui permettent d'utiliser la méthode du maximum de vraisemblance pour estimer les paramètres. Alors que les premières techniques impliquent une surestimation des effets et une sous-estimation des écart-types, notamment quand le réseau est dense (Lubbers et Snijders, 2007), le recours aux secondes ne présente pas ces limites notamment parce qu'elles permettent d'utiliser les propriétés mathématiques de ces distributions.

¹¹ Pour un réseau symétrique, binaire et sans possibilité de boucle (pas de possibilité de relation avec soi-même) de 30 nœuds, le nombre de possibilités est de $2^{\binom{30}{2}} = 8.87254302118661 \cdot 10^{43}$.

qui les construisent. Le corollaire est alors que ces sous-structures ont la même influence sur les acteurs. L'avantage est que les statistiques deviennent uniquement des comptabilisations de l'occurrence des configurations dans le réseau. Par exemple, si l'effet de réciprocité est intégré au modèle, sa statistique ($Z_{\mu\mu}$) correspond simplement au nombre de liens réciproques dans le réseau.

Cette hypothèse d'homogénéité n'est pas particulière aux ERGM, elle est commune à de nombreux modèles statistiques standards¹². Elle repose sur le postulat théorique selon lequel les acteurs du réseau étudié partagent des croyances et des normes de comportement. (Robins *et al.*, 2005). L'intégration d'effets exogènes, basés sur les attributs permet cependant de relâcher cette hypothèse d'homogénéité des configurations isomorphiques. Cette prise en compte d'effets exogènes (les attributs des acteurs) permet de saisir la double structuration de l'action : par les contraintes 1) d'ordre relationnel et 2) d'ordre sociodémographique. Un processus social spécifique peut alors être illustré en mixant attributs et structure dans le même esprit que les effets d'interaction entre variables dans les modèles plus traditionnels. Par exemple, Éloire (2010) montre ainsi une tendance à la solidarité indirecte au sein de l'un des différents types de marché des restaurants, au moyen d'une triade cyclique entre acteurs (en ayant contrôlé la cyclicité pour l'ensemble du graphe). À titre d'illustration, la troisième partie est consacrée à la mise en œuvre, pas à pas, de la modélisation ERGM sur des données empiriques.

<IT1>3. Application : le processus d'apprentissage dans un cabinet d'avocat

Notre objectif, dans cette partie, est d'étudier un réseau social sous l'angle de la triade. Pour cela, nous utilisons les données de Lazega (1992, 2001) concernant un cabinet d'avocats d'affaires américain afin d'y donner à voir l'existence d'un processus social d'apprentissage collectif, et la façon dont les sous-structures triadiques influent sur son déploiement au sein de l'organisation étudiée.

<IT2>Présentation du cas d'étude

Les 71 avocats du cabinet sont répartis entre trois bureaux dans différentes villes. Le cabinet est hiérarchisé : 36 avocats ont le statut d'associés (copropriétaires du cabinet) et 35 ont le statut de collaborateurs (avocats salariés). Les avocats sont spécialisés soit dans le conseil aux entreprises, soit dans le contentieux.

L'existence d'un processus d'apprentissage au sein du cabinet tient au fait que les besoins des clients dépassent souvent les compétences techniques de chaque avocat pris isolément. Ainsi, les relations informelles entre eux jouent un rôle essentiel en ce qu'elles permettent l'échange de ressources sociales, non facturées, au sein du cabinet. Un aspect de ce processus d'apprentissage est l'échange de connaissances spécifiques. Cependant, les avocats ont intérêt à agir sur un maximum de dossiers avec un minimum de collègues. Par conséquent, il s'agit de ne pas demander trop conseil à un autre avocat sur un dossier afin qu'il ne devienne partie-prenante du dossier. Les échanges de conseils constituent ainsi un indicateur de l'existence d'interdépendances épistémiques, économiques et symboliques illustrant la dimension informelle du processus d'apprentissage.

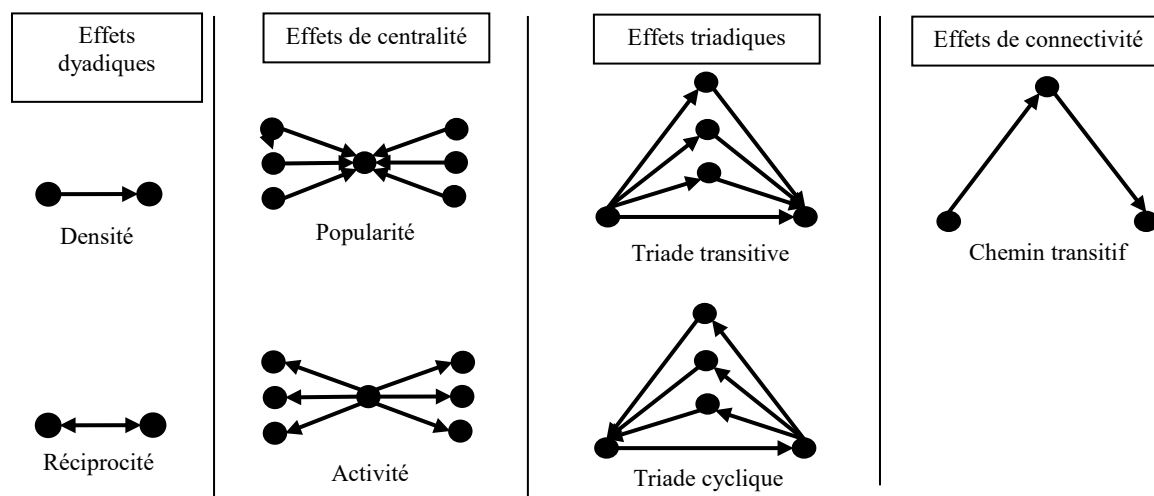
¹² Par exemple, les modèles linéaires généralisés supposent une homogénéité des effets quels que soient les acteurs. Si l'on cherche à estimer le revenu, sous réserve que l'on n'ait pas intégré de variables indicatrices pour certains groupes, on supposera par exemple que l'effet de du revenu du conjoint est le même pour tous les acteurs.

Ces échanges de conseils ont été reconstitués par Lazega (1992) au moyen d'un questionnaire sociométrique¹³. Les réponses des avocats permettent de reconstruire le réseau complet des échanges de conseils au sein du cabinet : il comprend 892 liens (densité : 17,9%) et 175 liens réciproques ; les avocats demandent conseil à 12,5 collègues en moyenne.

<IT2> Une analyse structurale sous l'angle des ERGM

Nous analysons le réseau en le décomposant en sous-structures locales qui peuvent revêtir des formes variées, dyadiques, triadiques, etc., que nous représentons sous forme de sous-graphes (voir Figure 2). L'objectif est de repérer quelles sont les configurations relationnelles qui pourraient résumer et répliquer le mieux ce réseau d'échanges de conseils. À chaque sous-structure particulière est affecté un paramètre statistique dont on peut apprécier l'importance pour la compréhension du réseau en le comparant au zéro (qui correspondrait à une sorte d'effet neutre ou aléatoire). Ainsi, si le paramètre afférant à une configuration est significativement positif (négatif) cela signifie que l'on observe plus (moins) de fois cette configuration que dans la moyenne de n réseaux simulés aléatoirement sous contrainte des autres paramètres. Cette configuration joue alors un rôle dans la structuration du système observé (voir Annexes 1 et 2).

Le Tableau 1 présente les résultats obtenus pour trois modèles ERGM estimés grâce au logiciel PNet. Chaque modèle correspond à un paramétrage différent. Nous présentons d'abord les résultats du Modèle 1 qui ne prend en compte que les effets structuraux (ou endogènes).



Note : Les effets de centralité et les effets triadiques sont d'ordre supérieur à leur équivalent markovien pour un facteur de correction (λ) de 5¹⁴.

Figure 2 : Effets structuraux mobilisés pour étudier le réseau entre avocats

¹³ « À qui parmi les personnes sur cette liste demandez-vous conseil lorsque vous êtes confronté à des problèmes d'ordre professionnel et que vous avez besoin d'un regard un peu extérieur au dossier ? Imaginez que vous avez des doutes sur la bonne manière de conduire une affaire importante et que vous voulez en discuter à fond avec quelqu'un (pas simplement demander un renseignement technique). Pourriez-vous indiquer sur cette liste les personnes que vous consultez dans ces cas-là ? » (Lazega, 1992).

¹⁴ Plus lambda est important et plus la valeur de l'effet se base sur le premier moment de la somme. Par conséquent plus lambda est important et plus le poids des 2-in-star dans la construction de l'effet de popularité est important, et donc plus l'effet correspond à son équivalent markovien.

	Effets	Type d'effet	Type de données	Modèle 1	Modèle 2	Modèle 3
EFFETS ENDOGENES	Densité	Contrôle	Dyadique	-1.640 (0.299) *	-4.137 (0.180) *	-2.948 (0.012) *
	Réciprocité	Structural	Dyadique	1.418 (0.196) *	1.567 (0.388) *	1.7729 (-0.05) *
	Popularité	Structural	Centralité	-0.207 (0.041) *	-	-0.091 (0.021)
	Activité	Structural	Centralité	-0.226 (0.044) *	-	-0.197 (0.017) *
	Triade Transitive	Structural	Triadique	0.806 (0.020) *	-	0.5242 (0.000) *
	Triade Cyclique	Structural	Triadique	-0.161 (0.018) *	-	-0.174 (-0.02) *
	Chemin transitif	Structural	Connectivité	-0.061 (0.005) *	-	-0.046 (0.026) *
EFFETS EXOGENES	Statuts	Interaction	Binaire	-	2.493 (0.225) *	1.4614 (-0.00) *
		Envoyeur	Binaire	-	-1.221 (0.215) *	-0.689 (0.059) *
		Receveur	Binaire	-	-0.175 (0.177)	-0.566 (-0.02) *
	Expérience	Envoyeur	Continue	-	-0.053 (0.009) *	-0.019 (-0.02) *
		Receveur	Continue	-	0.027 (0.007) *	0.0085 (0.008)
	Bureau	Interaction	Catégorielle	-	1.097 (0.132) *	0.3657 (-0.02) *
		Interaction Réciproque	Catégorielle	-	-0.916 (0.332) *	-0.829 (-0.06) *
	Formation	Interaction	Catégorielle	-	0.085 (0.106)	0.09 (-0.06)
	Spécialité	Interaction	Catégorielle	-	1.268 (0.127) *	0.7297 (0.017) *
		Interaction Réciproque	Catégorielle	-	-0.808 (0.277) *	-0.616 (-0.03) *
	Collaboration	Interaction	Var. Dyadique	-	2.166 (0.105) *	1.8828 (-0.00) *
Amitié	Interaction	Var. Dyadique	-	1.781 (0.133) *	1.4639 (0.031) *	

Note : La convergence et la significativité des estimateurs sont appréciées au moyen d'un test de Wald approché, dont le résultat est entre parenthèse. Si la valeur de l'estimateur est supérieure à 2 fois sa variance (approchée) alors l'effet de la configuration est significativement différent de 0 (Koskinen et Daraganova, 2013). Pour la procédure de validation des modèles, voir Annexe 3.

Tableau 1: Modèles ERGM estimés pour la présence de lien de conseil entre avocats

Concernant les effets dyadiques (densité et réciprocité), on constate que les échanges de conseils entre avocats ont tendance à être réciproques. Si l'avocat A demande conseil à l'avocat B, il y a de fortes chances pour que B sollicite A en retour. Comme nous l'avons évoqué, l'effet de réciprocité est un résultat fréquent en analyse de réseaux sociaux.

Concernant les effets de centralité, le fait que les paramètres de degré soient significativement négatifs implique qu'il n'y a pas d'individu extrêmement « populaire » (à qui on demanderait beaucoup conseil) ou extrêmement « actif » (qui demanderait beaucoup conseil). De fait, tous les avocats ont besoin de demander conseil à un moment ou à un autre. L'échange de conseils est une activité qui apparaît (relativement) équitablement distribuée et décentralisée.

Concernant les effets triadiques, il convient d'abord de les lire chacun isolément. Ce réseau se caractérise par un nombre de triades transitives plus important que ce que l'on aurait eu dans une distribution « aléatoire » ; contrairement aux triades cycliques qui elles sont moins fréquentes. Enfin, compte-tenu des autres effets présents dans le modèle (et donc notamment des triades transitives), il y a moins de chemins transitifs. Ce paramètre permet aussi de contrôler la corrélation entre la popularité et l'activité : les individus populaires ne sont généralement pas les individus actifs dans ce réseau.

Ensuite, il est intéressant de lire ces trois effets simultanément. L'articulation des paramètres de triades transitives et cycliques et de chemins transitifs peut être interprétée

comme une tendance hiérarchisée à la collaboration entre les avocats : il semble y avoir une pression sociale à la fermeture du chemin transitif par une relation directe, mais de manière transitive et non cyclique. Les résultats témoignent également de l'absence d'un effet d'intermédiation, tels ceux soulignés par Burt (1995) dans son approche des trous structuraux. Autrement dit, si A demande conseil à B et B demande conseil à C, généralement le chemin transitif se ferme parce que A demande également conseil à C, et non l'inverse. Le conseiller de mon conseiller est aussi mon conseiller. Le réseau a plutôt tendance à être aggloméré en triades transitives.

Ainsi, ce que nous apprend le Modèle 1, c'est que ce réseau se caractérise 1) par une tendance à la réciprocité, 2) par une participation de tous les avocats à l'échange de conseils, et 3) par une tendance à la fermeture des triades par transitivité et donc à une forme de hiérarchie. Mais cette analyse structurale (endogène) du réseau n'est qu'une première étape, car les attributs sociaux (exogènes) des avocats sont, eux aussi, essentiels pour comprendre le processus d'apprentissage au sein du cabinet.

<IT2> *Intégration des attributs : statuts, similarités et contextes*

Comme le souligne Blau (1964), les échanges sociaux sont organisés autour de logiques de statut, et de multiples processus d'homophilie permettent de contourner ces logiques statutaires. Demander conseil à quelqu'un, c'est lui reconnaître une forme de statut et d'autorité, et donner des conseils permet en retour d'entretenir un statut acquis, de sorte que les différences de statut constituent des frontières marquées. Pour contourner de telles logiques, les avocats peuvent s'appuyer sur d'autres registres relationnels, tels que l'amitié ou le fait de collaborer régulièrement avec une personne sur des dossiers. Ils peuvent aussi mobiliser des formes de ressemblance (homophilies) afin de détourner ces logiques de statut.

Le processus d'apprentissage est ainsi structuré par des logiques statutaires, que les avocats contournent par la mobilisation de critères d'identité communs, ou par l'existence de contextes relationnels multiplexes. Pour explorer cette idée, nous avons estimé un autre modèle ERGM, le Modèle 2, en intégrant uniquement des variables concernant des attributs des avocats ou de leurs relations (voir Encadré 2).

Encadré 2 : Attributs sociologiques et relationnels

1. Logiques statutaires :

Niveau hiérarchique des avocats dans le cabinet (variable binaire égale à 1 si l'avocat est associé)

2. Similarités permettant de contourner ces logiques de statut :

Bureau : variable catégorielle répertoriant le bureau de l'avocat (3 modalités).

Spécialité : variable catégorielle répertoriant la spécialité de l'avocat (2 modalités)

Formation : variable catégorielle répertoriant la formation de l'avocat (3 modalités)

3. Contextes favorisant ces relations :

Réseau de collaborations : variable dyadique égale à 1 si l'avocat i déclare avoir collaboré avec l'avocat j ?

Réseau d'amitié : variable dyadique égale à 1 si l'avocat i déclare être ami avec l'avocat j ?

4. Contrôler ces effets par l'expérience :

Expérience : variable continue répertoriant l'expérience de l'avocat (de 1 à 32 années)

Les effets liés au statut de l'avocat illustrent bien l'existence de logiques statutaires dans l'échange de conseils entre avocats : les associés interagissent plus entre eux que les collaborateurs (effet d'interaction positif) ; ils demandent aussi moins conseils que les collaborateurs (effet envoyeur négatif). Afin de savoir si les collaborateurs demandent plus de conseils aux associés que l'inverse, il est possible de croiser ces valeurs pour déterminer les *odds-ratio* des relations possibles selon le statut des avocats (voir Tableau 2). Les échanges de

conseils se font majoritairement entre avocats au statut identique, et les associés échangent beaucoup plus entre eux que les collaborateurs.

	Associés	Collaborateurs
Associés	2,995	0,295
Collaborateurs	0,839	1

Tableau 2 : Odds-Ratio pour le statut dans le Modèle 2 (voir annexe 4 pour la méthode de calcul)

Concernant les autres similarités, les effets sont significatifs et positifs, de sorte qu'il existe bien des critères qui permettent de contourner les logiques de statut : indépendamment du statut d'associé ou de collaborateur, les avocats peuvent demander conseil auprès d'autres avocats ayant la même spécialité ou travaillant au sein du même bureau. Les résultats traduisent ainsi une certaine organisation du travail à la fois géographique et sectorielle. Cependant, les échanges entre avocats d'un même bureau et entre avocats d'une même spécialité ont tendance à être moins réciproques que les échanges hétérophiles, ce qui traduit, là encore, une certaine forme de hiérarchie. Enfin, la présence d'un lien d'amitié ou de collaboration sur un dossier entre les deux avocats s'avère être un contexte favorable à l'échange de conseils entre eux.

Afin d'éviter l'*erreur structurale* lorsqu'on interprète les effets des attributs, il est nécessaire de les estimer simultanément avec les effets structuraux endogènes. Le Modèle 3 conduit alors à nuancer le résultat selon lequel les associés échangent plus entre eux que les collaborateurs. Le calcul des *odds-ratio* pour ces nouvelles valeurs (voir Tableau 3) révèle une différence beaucoup plus faible. Une part importante des liens homophiles entre associés s'explique donc aussi par des effets de centralité et de fermeture triadiques.

	Associés	Collaborateurs
Associés	1,229	0,502
Collaborateurs	0,568	1

Tableau 3 : Odds-Ratio pour le statut dans le Modèle 3

En outre, les résultats du Modèle 3 nuancent ceux du Modèle 2 : les avocats les plus expérimentés ne sont pas nécessairement plus populaires que les autres. Cette analyse montre la pluralité des facteurs expliquant les relations de conseils entre avocats, et met l'accent sur l'importance des effets triadiques, de centralité et de fermeture notamment, au sein du processus d'apprentissage. Cela permet ainsi d'éviter une interprétation trop centrée sur les seuls phénomènes d'homophilie, et de montrer qu'à l'échelle locale, les processus sociaux se déploient de manière complexe, faisant intervenir certes les attributs sociaux, mais aussi les effets endogènes liés aux voisinages relationnels dans lesquels les dyades sont insérées. Les modèles ERGM apporte ainsi une compréhension fine de la structure de ce réseau et cet exemple montre l'intérêt de ces modèles en comparaison d'autres outils d'analyse de réseaux notamment ceux se limitant à une analyse dyadique. D'une part ils permettent de comprendre les mécanismes relationnels dyadiques et triadiques qui conduisent à produire la structure globale du réseau telle qu'on peut l'observer, d'autre part ils permettent de contrôler les effets de structure et donc de montrer de manière robuste l'effet des attributs dans la formation des relations.

<IT1>4. Des exemples récents de travaux mobilisant des ERGM

Un nombre croissant de travaux mobilisent les ERGM sur des terrains d'études très divers. Ainsi, de nombreuses études ont été menées au sein d'écoles, collèges ou lycées afin d'étudier les types de relations que nouent les élèves entre eux. Certains auteurs soulignent ainsi les processus de ségrégation et de discrimination liés à l'ethnicité au sein des écoles (Goodreau et al., 2009 ; Vermeij et al., 2009). Mais, comme nous l'avons dit, c'est à l'échelle de la triade que des hypothèses originales peuvent être formulées et testées. L'étude de Huitsing et Veenstra (2012) constitue ainsi un bon exemple de l'apport des ERGM, en s'intéressant à deux types de relations dans des classes d'écoles élémentaires aux Pays-Bas, d'une part des relations de harcèlement et d'insultes entre les élèves, et d'autre part des relations d'aide ou de défense lorsqu'une personne est harcelée, insultée ou intimidée. À l'aide d'un ERGM, ils analysent simultanément ces deux types de relations et montrent que les élèves ont tendance à s'entraider lorsqu'ils sont intimidés par les mêmes personnes, et parallèlement que les élèves intimidant les mêmes personnes s'aident mutuellement. Ces liens « négatifs » sont souvent à l'origine de la formation des groupes sociaux : c'est en partie parce que l'on entretient des relations de défiance envers les mêmes personnes que l'on noue des relations amicales.

D'autres travaux s'intéressent aux relations inter-organisationnelles. C'est le cas de Pallotti et al. (2013) qui reconstituent les échanges de patients entre les hôpitaux du Latium en Italie et qui montrent que la structure du réseau dépend autant d'effets de proximité et de complémentarité entre les hôpitaux que de logiques structurales. Les liens d'échanges de patients entre les hôpitaux ne sont pas indépendants les uns des autres et les auteurs révèlent des mécanismes de fermeture triadique qui ne s'expliquent ni par la proximité des hôpitaux ni par la complémentarité de leurs expertises. Éloire (2010) utilise les ERGM pour révéler des formes de solidarité entre les restaurateurs de la région Lilloise. Ceux-ci s'échangent des services, se renvoient des clients et s'échangent des « bons plans » pour leur approvisionnement en denrées. Piña-Stranger et Lazega (2011) étudient les échanges de conseils informels entre les entrepreneurs français en biotechnologie. À l'échelle triadique, ils soulignent le rôle des investisseurs dans le processus d'apprentissage collectif entre ces entrepreneurs : ceux-ci jouent bien souvent un rôle d'entremetteur ; on peut ainsi observer de multiples triades incluant des capitaux-risqueurs et des entrepreneurs. Favre (2014) utilise les ERGM pour analyser conjointement un réseau de discussion et un réseau de signature de contrats commerciaux entre les participants d'un salon de distribution de programmes de télévision en Afrique sub-saharienne. Il montre que les participants les plus réguliers du salon, les « habitués », parviennent à maintenir une position dominante dans ce marché en échangeant des informations avec leurs concurrents directs et notamment ceux qui concluent des contrats avec les mêmes clients qu'eux, là où les nouveaux entrants, les « novices » sont forcés de se limiter à des discussions avec leurs partenaires commerciaux.

De nombreux travaux appliquent également ces modèles aux réseaux de relations informelles au sein des organisations, afin de révéler les structures concrètes de relations de travail au-delà de l'organigramme de l'entreprise. Rank et al. (2010) étudient conjointement des relations d'amitié, d'échanges d'information et de collaboration entre les membres de deux entreprises allemandes. À l'aide des ERGM, ils étudient les différentes triades multiplexes et montrent qu'il existe une tendance à la fermeture triadique de ces relations, mais que les triades cycliques, c'est-à-dire les formes d'échange généralisé, sont sous-représentées. Ils confirment ainsi les travaux menés plus tôt par Lazega et Pattison (1999) révélant les mêmes mécanismes relationnels dans le cabinet d'avocat précédemment évoqué.

<IT1>5. Perspectives et limites

En ce qui concerne le développement des ERGM, plusieurs perspectives récentes peuvent être soulignées. Premièrement, des raffinements au modèle général ont été apportés afin de pouvoir étudier les phénomènes d'influence ou de diffusion sociales à partir de données de réseau. Il s'agit des *Autologistic Actor Attribute Models* ou ALAAM (Daraganova et Robins, 2013). Dans le cadre de ces modèles, les liens sont exogènes et les attributs endogènes, et on cherche à modéliser le comportement des acteurs en se demandant : quelles configurations de relations sociales sont associées avec quels types d'opinions partagées ou de comportements similaires ? Des travaux en cours ambitionnent d'articuler ces modèles ALAAM aux modèles ERGM afin d'étudier simultanément la sélection sociale (ERGM) et l'influence sociale (ALAAM).

La deuxième perspective concerne les analyses de réseaux dits multiniveaux qui sont beaucoup développées depuis quelques années (Lazega et Snijders, 2016). Ces réseaux se caractérisent par la présence de plusieurs niveaux d'acteurs, par exemple des individus travaillant dans des organisations, et donc de trois réseaux : un réseau entre individus, un réseau entre organisations et un réseau d'affiliation reliant les deux niveaux) (Brailly et Lazega, 2012 ; Brailly et al., 2016 ; Favre et al., 2016). Afin d'étudier ce type de données, des modèles ERGM spécifiques ont été développés (Wang et al., 2013) qui permettent notamment d'étudier des effets de transitivité d'un niveau à l'autre : par exemple, des entreprises populaires dans leur système relationnel parce que composées d'individus eux-mêmes populaires dans leur système.

D'autres travaux en cours portent sur l'assouplissement de l'hypothèse d'homogénéité des configurations isomorphiques dont nous avons parlé plus haut. Certains proposent de permettre aux estimateurs de différer en partie selon un attribut des acteurs (Pina-Stranger et Lazega, 2011 ; Brailly, 2014). Il ne s'agit pas d'une remise en cause l'hypothèse d'homogénéité des configurations, nécessaire à la procédure d'estimation, mais plutôt d'un relâchement de celle-ci à la marge, réalisable au moyen des logiciels développés pour l'analyse des réseaux multiniveaux. D'autres, proposent de faire différer les estimateurs selon des zones du graphe : cœur/périphérie ou groupes structurellement équivalents (Schweinberger et Handcock, 2009).

En ce qui concerne les limites de ces voies de recherche, l'une des principales concerne le nombre de paramètres nécessaires qui, quel que soit le type de données devient très important pour ajuster le modèle, du fait notamment de l'introduction d'effets d'interaction. Il est alors fréquent de voir des modèles comprenant plus de quarante effets pour tester des hypothèses qui pourraient être étudiées avec des modélisations plus simples. Un enjeu de cette contribution consiste justement à insister sur les fondements théoriques implicites de tels modèles, et à les ancrer dans la réflexion sociologique. Si ces modèles enrichissent souvent l'analyse en révélant des phénomènes invisibles avec d'autres outils, il ne s'agit pas de tomber dans une dérive quantitativiste. D'un côté, la parcimonie dans l'estimation des modèles doit être de rigueur. De l'autre, il est nécessaire de compléter ces modélisations par des analyses descriptives telles que les modèles de blocks (pour une présentation de l'ensemble de ces outils voir Doreian et al. 2005). Il est toujours nécessaire de les articuler à une ethnographie poussée, et de revenir aux activités concrètes des acteurs. Enfin, il est toujours souhaitable, quand cela est possible, de les articuler à des modèles plus dynamiques tels que les modélisations de type SIENA (Snijders et al. 2010).

Une autre limite concerne la puissance computationnelle requise par les simulations nécessaires à l'estimation des modèles ERGM, qui ne permettent pas réellement aujourd'hui d'analyser des réseaux de plus cinq cents acteurs. Par exemple le dernier modèle du tableau 3

a nécessité plus de deux heures de calcul informatique pour un réseau composé de 71 individus. Il faut ajouter que la démarche d'estimation est inductive, et nécessite d'alterner entre l'ajout de paramètres et l'étude de la qualité d'ajustement du modèle. L'ensemble de ces étapes peut donc prendre beaucoup de temps que seules l'expérience et la familiarité avec des données de réseau permet de minimiser.

Comparée à d'autres méthodes statistiques, une autre limite concerne l'impossibilité de comparer les valeurs des estimateurs de deux modèles ERGM, quand bien même ils concerneraient les mêmes données. En effet, l'estimation des paramètres étant basée sur la simulation de réseaux selon des caractéristiques définies à l'avance mais faisant intervenir un facteur aléatoire, il en résulte que les facteurs de corrections sont toujours différents d'un modèle à l'autre. Il est uniquement possible de comparer la significativité de ces estimateurs entre différents modèles. Enfin, une dernière limite concerne la sensibilité des modèles ERGM aux non-réponses, pouvant mener à des sous ou surestimations de certains effets (Robins et *al.*, 2004). De sorte que le chercheur doit garder à l'esprit que ses résultats sont uniquement valables pour la population des répondants.

<IT1>Conclusion

Il est souvent considéré en sociologie que les opinions ou comportements des individus sont déterminés par les structures sociales. L'analyse des réseaux sociaux propose d'affiner cette logique catégorielle en considérant que les normes et les comportements sont dépendent des réseaux de relations des individus (Degenne et Forsé, 1994). L'analyse à partir des ERGM permet de combiner ces deux approches, car elle prend à la fois en compte ces logiques catégorielles, « exogènes » au réseau, et les effets « endogènes », les mécanismes purement relationnels entre individus, pour comprendre de manière fine l'effet des structures sociales sur les relations, et l'effet des relations sur les structures sociales. En mesurant les effets triadiques dans les réseaux sociaux, les modèles ERGM formalisent les intuitions de Simmel et d'autres auteurs qui considéraient que les structures relationnelles pouvaient être à l'origine de phénomènes sociaux à l'échelle d'un collectif.

Encadré 4 : Pour aller plus loin...

De nombreux articles sur les modèles ERGM ont été publiés en anglais, qu'ils concernent des applications ou des développements statistiques. Pour le chercheur qui voudrait aller plus loin, l'ouvrage de référence est à ce jour celui codirigé par Lusher, Koskinen et Robins (2013) qui fait le point sur les fondements statistiques des modèles ERGM, avec une première partie correspondant à une introduction accessible aux non-initiés (sans aucune équation) et une seconde partie plus exhaustive d'un point de vue statistique, donc réservée à un public averti ; et présente huit cas d'études auxquels cette méthode est appliquée.

L'ouvrage reste néanmoins principalement méthodologique, on trouvera une présentation et une réflexion plus théorique sur les ERGM dans l'article de Robins, Pattison et Woolcock (2005).

Les principaux logiciels permettant de réaliser des ERGM sont :

1. Le logiciel PNet (Wang *et al.*, 2006) : <http://www.swinburne.edu.au/fbl/research/transformational-innovation/our-research/MelNet-social-network-group/PNet-software/index.html>
2. Le paquet ERGM pour le logiciel R (Hunter *et al.*, 2008) : http://statnet.csde.washington.edu/workshops/SUNBELT/EUSN/ergm/ergm_tutorial.html

Sur ces deux sites, des tutoriels précis, ainsi que des jeux de données et des cours permettent d'apprendre progressivement à estimer les ERGM. Le site personnel de Tom Snijders fournit aussi une documentation précieuse (<https://www.stats.ox.ac.uk/~snijders/siena/>).

Enfin, il existe plusieurs formations, chaque année, dédiée à ces modèles. Depuis 2013, Julien Brailly, Scott Viallet-Thévenin, Guillaume Favre et Emmanuel Lazega organisent à SciencesPo Paris, au Centre de Sociologie des Organisations, une formation en langue française. De nombreuses formations sont disponibles en anglais lors d'écoles d'été ou encore durant la conférence annuelle des analystes de réseaux sociaux, la Sunbelt (<http://insna.org/sunbelt2016/>).

<BIBLIOGRAPHIE>Références bibliographiques

- Besag J., 1974, « Spatial interaction and the statistical analysis of lattice systems », *Journal of the Royal Statistical Society. Series B (Methodological)*, pp.192-236.
- Blau P. M., 1964, *Exchange and power in social life*, Transaction Publishers.
- Bourdieu P., 2012, *Sur l'Etat. Cours au Collège de France 1989-1992*, Seuil, Raisons d'Agir.
- Brailly J., Lazega E., 2012, « Diversité des approches de modélisation statistique en analyse de réseaux sociaux multiniveaux », *Mathématiques et sciences humaines*, 198, pp. 5-32.
- Brailly J., 2014, *Coopérer pour résister : Interactions marchandes et réseaux multiniveaux dans un salon d'échange de programmes de télévision en Europe Centrale et Orientale*, Thèse de doctorat, Université Paris–Dauphine, 13 novembre.
- Brailly J., Favre G., Chatellet J., Lazega E., 2016, « Embeddedness as a Multilevel Problem. A Case Study in Economic Sociology », *Social Networks*, 44(1), pp. 319-333
- Burt R., 1995, « Le capital social, les trous structuraux et l'entrepreneur », *Revue Française de Sociologie*, 36, pp. 599-628.
- Caplow T., 1968 [1971], *Deux contre un. Les coalitions dans les triades*, Paris, Armand Colin.
- Degenne A., Forsé M., 1994 [2004], *Les réseaux sociaux*, Paris : Armand Colin.
- Doreian, P., Batagelj, V., Ferligoj, A., 2005, *Generalized blockmodeling*, Cambridge University Press.
- Éloire F., 2010, « Une approche sociologique de la concurrence sur un marché. Le cas des restaurateurs lillois », *Revue française de sociologie*, 2010/3 (Vol. 51), pp. 481-517.
- Éloire F., Penalva-Icher E., Lazega E., 2011- « Application de l'analyse des réseaux complets à l'échelle interorganisationnelle », *Terrains & travaux*, 2, pp. 77-98.
- Erdős P., Rényi A., 1959, « On random graphs », *Publicationes Mathematicae Debrecen*, 6, pp. 290-297.
- Favre G., 2014, *Des rencontres dans la mondialisation: réseaux et apprentissages dans un salon de distribution de programmes de télévision en Afrique sub-saharienne*, Thèse de doctorat, Université Paris Dauphine, 2 décembre.
- Favre, G., Brailly, J., Chatellet, J., Lazega, E., 2016, « Inter-organizational network influence on long-term and short-term inter-individual relationships: The case of a trade fair for TV programs distribution in sub-Saharan Africa », in Lazega E., Snijders T.A.B. *Multilevel Network Analysis for the Social Sciences*, Springer International Publishing, pp. 295-314.
- Frank O., Strauss D., 1986, « Markov graphs », *Journal of the American Statistical Association*, 81, pp. 832-842.
- Goodreau S., Kitts J.A., Morris M., 2009, « Birds of a Feather, or Friend of a Friend? Using Exponential Random Graph Models to Investigate Adolescent Social Networks », *Demography*, 46-1, pp.103-125.
- Granovetter M., 1973, « The strength of weak ties », *American Journal of Sociology*, 78 (6), pp.1360-1380. Traduction en français in. *Le marché autrement*, 2000, Paris, Desclee de Brouwer.
- Hammersley J. M., Clifford P., 1971, « *Markov fields on finite graphs and lattices* », document de travail
- Holland P.W., Leinhardt S., 1976, « Local Structure in Social Networks », *Sociological Methodology*, 7, pp. 1-45.

- Huitsing G., Veenstra R., 2012, « Bullying in classrooms: participant roles from a social network perspective », *Aggressive behavior*, 38-6, pp.494-509.
- Hunter D. R., Handcock M. S., Butts C. T., Goodreau S. M., Morris M., 2008, « ergm: A package to fit, simulate and diagnose exponential-family models for networks », *Journal of Statistical Software*, 24.
- Koskinen, J. H., Robins, G. L., Pattison, P. E., 2010, « Analysing exponential random graph (p-star) models with missing data using Bayesian data augmentation », *Statistical Methodology*, 7(3), 366-384.
- Koskinen J., Daraganova G., 2013, « Dependence Graphs and Sufficient statistics », in Lusher D., Koskinen J., Robins G.L., (éds.), *Exponential Random Graph Models for Social Networks: Theory, Methods and Applications*, New York, Cambridge University Press, pp. 77-90.
- Laumann E. O., Marsden P. V., Prensky D., 1989, « The boundary specification problem in network analysis », *Research methods in social network analysis*, 61, pp. 87.
- Lazega E., Jourda M.-T., Mounier L., Stofer R., 2007, « Des poissons et des mares : l'analyse de réseaux multiniveaux », *Revue Française de Sociologie*, 48, pp. 93-131.
- Lazega E., 1998 [2014], *Réseaux sociaux et structures relationnelles*, Paris, Presses Universitaires de France, coll. Que-sais-je?.
- Lazega E., Pattison P., 1999, « Multiplexity, generalized exchange and cooperation in organizations: a case study », *Social networks*, 21, pp. 67-90.
- Lazega E., 2001, *The Collegial phenomenon*, Oxford, Oxford University Press.
- Lévi-Strauss C., 1967, *Les structures élémentaires de la parenté*, Paris, La Haye, Mouton Editeur, éd. La Maison des Sciences de l'Homme.
- Lazega, E., Snijders, T. A.B., 2015, *Multilevel Network Analysis for the Social Sciences*, Springer International Publishing,
- Lubbers, M. J., Snijders, T. A.B., 2007, A comparison of various approaches to the exponential random graph model: A reanalysis of 102 student networks in school classes. *Social Networks*, 29(4), 489-507.
- Lusher D., Koskinen J., Robins, G. (éds), 2013, *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*, New York, Cambridge University Press, Structural Analysis in the Social Sciences Series.
- Malinowski B., 1922 [1989], *Les Argonautes du Pacifique occidental*, Paris, Gallimard.
- Mauss M., 1950 [2004], « Essai sur le don », in. *Sociologie et anthropologie*, Paris, PUF, Quadrige.
- Pallotti F., Lomi A., Mascia D., 2013, « From network ties to network structures: Exponential Random Graph Models of interorganizational relations », *Quality & Quantity*, 47, pp.1665-1685.
- Pattison P.E., Robins G., 2002, « Neighbourhood-based models for social networks », *Sociological Methodology* 32, pp. 301-337.
- Piña-Stranger A., Lazega E., 2011, « Bringing personalized ties back in: their added value for biotech entrepreneurs and venture capitalists interorganizational networks », *The Sociological Quarterly*, 52, pp. 268-292.
- Rank O.N., Robins G.L., Pattison P.E., 2010, « Structural Logic of Intraorganizational Networks », *Organization Science*, 21-3, pp.745-764.

- Robins, G., Pattison, P., Woolcock, J., 2004, « Missing data in networks: exponential random graph (p*) models for networks with non-respondents », *Social Networks*, 26(3), 257-283.
- Robins G.L., Pattison P.E., Woolcock J., 2005, « Social networks and small worlds », *American Journal of Sociology*, 110, pp. 894-936.
- Robins G.L., Pattison P.E., Kalish Y., Lusher D., 2007, « An introduction to exponential random graph (p*) models for social networks », *Social Networks*, 29, pp. 173-191.
- Robins G.L., Snijders T.A.B., Wang P., Handcock M., Pattison P.E., 2007, « Recent developments in exponential random graph (p*) models for social networks », *Social Networks*, 29, pp. 192-215.
- Robins G.L., Lusher D., 2013, « What are exponential random graph models (ERGM)? », in Lusher D., Koskinen J., Robins G., (éds.), *Exponential Random Graph Models for Social Networks: Theory, Methods and Applications*, New York, Cambridge University Press.
- Robinson W.S., 1950, « Ecological correlations and the behaviour of individuals », *American Sociological Review*, 15, p. 351-357.
- Simmel G., 1908 [1999], *Sociologie : études sur les formes de socialisation*, PUF, Sociologie.
- Snijders T.A.B., 2002, « Markov Chain Monte Carlo estimation of exponential random graph models », *Journal of Social Structure*, 3(21), pp. 187-209.
- Snijders T.A.B., Pattison P.E., Robins G.L., Handcock M., 2006, « New specifications for exponential random graph models », *Sociological Methodology*, 36, pp. 99-153.
- Snijders T.A.B., 2011, « Statistical models for social networks », *Annual Review of Sociology*, 37, pp. 129-159.
- Snijders T.A.B., Van de Bunt G.G., Steglich C.E., 2010, « Introduction to Stochastic Actor-based Models for Network Dynamics », *Social Networks*, 32-1, pp. 44-60.
- Steiner P., 2009, *La transplantation d'organes, un commerce entre humains*, Paris, NRF Gallimard.
- Vermeij L., Van Duijn M.A.J., Baerveldt C., 2009, « Ethnic segregation in context: Social discrimination among native Dutch pupils and their ethnic minority classmates », *Social Networks*, 31-4, pp. 230-239.
- Wang P., Robins G.L., Pattison P.E., 2006, *Pnet: a program for the simulation and estimation of exponential random graph models*, <http://sna.unimelb.edu.au/PNet>.
- Wang P., Robins G.L., Pattison P.E., Lazega E., 2013, « Exponential random graph models for multilevel networks », *Social Networks*, 35, pp. 96-115.
- Wasserman, S., Faust, K., 1994, *Social network analysis: Methods and applications* (Vol. 8). Cambridge University Press.
- White H.C., Boorman S.A., Breiger R. L., 1976, « Social structure from multiple networks. I. Blockmodels of roles and positions », *American journal of sociology*, pp. 730-780.

Annexe 1 : Procédure d'estimation des modèles ERGM

D'un point de vue algorithmique et computationnel, l'estimation est réalisée au moyen d'une procédure appelée *Monte-Carlo Markov Chain Maximum Likelihood Estimation* (MCMCMLE) qui permet de déterminer la distribution des paramètres du modèle en se basant sur des graphes générés aléatoirement selon les configurations spécifiées dans le modèle. La procédure d'estimation consiste à sélectionner aléatoirement des liens dans les graphes aléatoires, et à changer uniquement le lien et mettre à jour les estimateurs si cela se rapproche de la distribution observée empiriquement. Tant que les paramètres estimés par la simulation ne convergent et ne stabilisent pas, ils sont révisés (Snijders, 2002). Cette procédure fournit ainsi une distribution de probabilités pour l'ensemble des paramètres sélectionnés, au départ, par le chercheur lui-même au moment du paramétrage du modèle (en fonction de ses hypothèses de recherche).

La procédure d'estimation est réalisée via l'échantillonneur Metropolis qui consiste à sélectionner aléatoirement de manière séquentielle m liens. À chaque séquence, le lien est changé et les configurations recalculées. Ce changement est gardé si cela rapproche de la distribution de probabilités ciblée.

Annexe 2 : Hiérarchie des effets

Il est possible de distinguer quatre types d'effets endogènes (voir figure 2) : les effets dyadiques (densité et réciprocité), les effets de centralité (activité et popularité), les effets de fermeture triadique (propension des individus à agir dans des groupes), les effets de connectivité (structures triadiques non fermées). Les inférences sur un effet d'ordre supérieur sont conditionnées par les effets d'ordre inférieur qui doivent donc être présents dans le modèle. Par conséquent, il ne serait pas raisonnable de modéliser un réseau en incluant uniquement des effets triadiques. Il faut au préalable inclure la densité du réseau (plus un réseau est dense, plus il y a de chance d'avoir des triades) des effets de distribution de degré (plus il y a des acteurs populaires et plus liens ont tendances à se concentrer dans certaines zones du réseau), des effets de connectivité tel que les chaînes transitives (plus il y a de pas-2 et plus il y a des chances d'avoir des triades). C'est seulement une fois que ces différents effets sont pris en compte qu'il est possible d'intégrer les effets triadiques sans risquer de tomber dans l'erreur structurale. Il en va de même pour les effets d'attributs exogènes. Si l'on veut étudier la solidarité indirecte entre les acteurs de type A, il est nécessaire d'avoir dans le modèle le paramètre triade cyclique pour l'ensemble du réseau et celui d'homophilie entre les acteurs de type A, puis en plus, pour les acteurs de type A.

Annexe 3 : Validation des modèles et Goodness of Fit

La validation des modèles se fait en deux étapes :

Etape 1

La convergence et la significativité des estimateurs sont appréciées au moyen d'un test de Wald (approché). Si la valeur de l'estimateur est supérieure à deux fois sa variance (approchée) alors l'effet de la configuration est significativement différent de zéro (Koskinen et Daraganova, 2013).

Etape 2

Cependant dans la mesure où l'estimation est totalement dépendante des configurations spécifiées dans le modèle, il est intéressant de valider le modèle non par rapport à un autre modèle réalisé sur les mêmes données comme c'est le cas pour les tests, mais en regardant la qualité d'ajustement générale du modèle par rapport au réseau empirique. L'idée est alors de générer de manière aléatoire un grand nombre de réseaux selon les paramètres estimés et de regarder si les réseaux générés ont les mêmes caractéristiques que le réseau observé empiriquement.

L'intérêt est alors de vérifier que les caractéristiques modélisées s'ajustent bien (ce qui est logique) mais aussi l'ensemble de configurations au voisinage, les distributions de degré ainsi que plusieurs mesures à l'échelle du réseau dans son ensemble, tels que le *coefficient d'agglomération*.

Le modèle choisi doit donc reproduire fidèlement les caractéristiques sélectionnées par le chercheur, mais aussi un nombre important d'indicateurs n'étant pas explicitement dans le modèle.

Pour les modèles du Tableau 1, les résultats de ce deuxième niveau de validation du modèle sont présentés dans la suite de cette annexe (plus de 150 caractéristiques locales ou globales du réseau sont vérifiées, les effets inclus dans les modèles sont en gras). Par convention le modèle réplique bien les caractéristiques du réseau observé si le *t-ratio* est inférieur à 2 pour tous les effets et 0.5 pour les effets inclus dans le modèle.

$$t - ratio = \frac{\text{Observé} - \text{Moyenne des simulations}}{\text{Ecart - type des simulations}}$$

Tableau annexe 3 : Goodness of fit pour le modèle 3

Effets	Observé	Moyenne des simulations	Ecart-type des simulations	t-ratio
Densité	892	889,5	29,0	0,085
Réciprocité	175	173,9	10,8	0,105
2-in-star	7627	7562,4	534,9	0,121
2-out-star	6961	6718,9	445,9	0,543
3-in-star	50761	49666,9	5750,2	0,19
3-out-star	41192	37428,2	3979,7	0,946
Chemin transitif	11447	11463,6	759,3	-0,022
T1	144	159,1	26,9	-0,562
T2	1306	1342,6	186,8	-0,196
T3	1888	1874,8	223,7	0,059
T4	1407	1356,9	140,7	0,356
T5	1242	1278,4	130,9	-0,278
T6	1325	1261,3	161,9	0,393
T7	5873	5730,1	527,6	0,271
T8	5321	5194,4	472,7	0,268
T9(030T)	5075	4985,9	397,6	0,224
T10(030C)	880	872,4	93,6	0,081
Sink	1	0,8	0,8	0,309
Source	1	1,1	1,0	-0,115
Isolates	0	0,0	0,1	-0,138
Popularité (Lambda = 5)	3022,7	3008,2	136,9	0,106
Activité (Lambda = 5)	2951,8	2942,1	134,9	0,072
AinS(2,00)	1516,0	1510,8	57,8	0,09
AoutS(2,00)	1508,6	1506,4	57,9	0,039
AinIout-star(2,00)	1695,1	1698,4	61,5	-0,054
IinAout-star(2,00)	1713,3	1705,0	60,6	0,137
AinAout-star(2,00)	259,5	257,8	4,9	0,357
AT-T(5,00)	2818,2	2801,7	159,3	0,104
AT-C(5,00)	1644,5	1623,0	132,8	0,162
AT-D(2,00)	1455,9	1436,4	65,8	0,297
AT-U(2,00)	1532,5	1494,8	67,3	0,561
AT-TD(2,00)	1510,0	1489,5	65,0	0,315
AT-TU(2,00)	1548,3	1518,7	65,8	0,449
AT-DU(2,00)	1494,2	1465,6	63,8	0,45
AT-TDU(2,00)	1517,5	1491,3	64,3	0,408
A2P-T(5,00)	7465,7	7422,4	368,2	0,118
A2P-D(2,00)	2766,9	2683,3	123,8	0,675
A2P-U(2,00)	3078,1	3070,3	137,0	0,057
A2P-TD(2,00)	3805,3	3721,9	149,1	0,56
A2P-TU(2,00)	3960,9	3915,4	146,4	0,311
A2P-DU(2,00)	2922,5	2876,8	104,1	0,439

A2P-TDU(2,00)	3562,9	3504,7	126,5	0,46
status_1_interaction	395,000	395,897	19,146	-0,047
status_2_interaction	196,000	196,350	13,238	-0,026
gender_1_interaction	579,000	556,643	21,514	1,039
status_1_sender	454,000	452,683	19,417	0,068
status_2_sender	438,000	436,839	21,753	0,053
gender_1_sender	680,000	686,434	23,304	-0,276
status_1_receiver	637,000	636,386	26,655	0,023
status_2_receiver	255,000	253,136	15,452	0,121
gender_1_receiver	743,000	714,402	26,891	1,063
status_1_interaction_reciprocity	106,000	107,265	8,776	-0,144
status_2_interaction_reciprocity	37,000	36,410	5,583	0,106
gender_1_interaction_reciprocity	119,000	115,285	8,741	0,425
years_Sender	8516,000	8493,582	332,957	0,067
age_Sender	35829	35926,701	1172,867	-0,083
years_Receiver	12520	12479,303	503,58	0,081
age_Receiver	39678	39428,001	1337,145	0,187
office_matching	700	697,544	22,752	0,108
Law_matching	322	320,3	14,429	0,118
speciality_matching	661	657,566	22,757	0,151
office_matching_reciprocity	145	143,576	9,926	0,143
Law_matching_reciprocity	63	64,913	6,183	-0,309
speciality_matching_reciprocity	132	130,82	9,457	0,125
AMITIE_arc	358	357,889	10,898	0,01
COWORK_arc	582	580,238	15,393	0,114
Std Dev in-degree dist	8,4	8,319	0,45	0,181
Skew in-degree dist	0,576	0,512	0,137	0,473
Std Dev out-degree dist	7,179	6,72	0,374	1,226
Skew out-degree dist	0,497	0,286	0,171	1,232
CorrCoef in-out-degree dists	0,138	0,165	0,063	-0,431
Global Clustering Cto	0,365	0,371	0,015	-0,439
Global Clustering Cti	0,333	0,33	0,011	0,288
Global Clustering Ctm	0,443	0,435	0,013	0,662
Global Clustering Ccm	0,231	0,228	0,015	0,175
Global Clustering AKC-T	0,323	0,324	0,011	-0,121
Global Clustering AKC-D	0,263	0,268	0,009	-0,545
Global Clustering AKC-U	0,249	0,244	0,007	0,831
Global Clustering AKC-C	0,21	0,208	0,009	0,259

Annexe 4 : Odds-Ratio pour le statut

	Associés	Collaborateurs
Associés	$\exp(\text{effet envoyeur} + \text{effet receveur} + \text{effet similarité})$	$\exp(\text{effet envoyeur})$
Collaborateurs	$\exp(\text{effet receveur})$	1