



HAL
open science

Modélisation hybride de réseaux dans un "champ criminel" : contribution des sciences sociales et d'outils logiciels au renseignement criminel

Maud Benichou Duhil de Benaze

► To cite this version:

Maud Benichou Duhil de Benaze. Modélisation hybride de réseaux dans un "champ criminel" : contribution des sciences sociales et d'outils logiciels au renseignement criminel. Histoire. Université Michel de Montaigne - Bordeaux III, 2023. Français. NNT : 2023BOR30002 . tel-04011375

HAL Id: tel-04011375

<https://theses.hal.science/tel-04011375>

Submitted on 2 Mar 2023

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

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

Université Bordeaux Montaigne

École Doctorale Montaigne Humanités (ED 480)

Laboratoire : Institut de recherches Montesquieu (IRM) de l'Université de Bordeaux

Thèse de doctorat en « histoire contemporaine »

Modélisation hybride de réseaux dans un « champ criminel » : contribution des sciences sociales et d'outils logiciels au renseignement criminel.

Présentée et soutenue publiquement le 6 janvier 2023 par

Maud BENICHO DUHIL de BENAZE

Sous la direction de Sébastien-Yves Laurent

Membres du jury :

Lt-Col Jérôme BARLATIER	Chef de la division renseignement, SCRC Gendarmerie nationale.
Gautier CLOIX	Directeur général Palantir France, invité
Xavier CRETTEZ	Professeur des Universités, Université de Versailles, Président du jury.
Christophe FAGOT	Responsable R&D, Intactile Design.
Sébastien-Yves LAURENT	Professeur des Universités, Université de Bordeaux, directeur de la thèse
Guy MELANCON	Professeur des Universités, Université de Bordeaux.
Bruno PINAUD	Maître de conférences HDR, Université de Bordeaux.
Olivier RIBAUX	Professeur ordinaire, Université de Lausanne, rapporteur.
Gilles VENTURINI	Professeur des Universités, Université de Tours, rapporteur.

Remerciements

Je souhaite tout d'abord exprimer mes remerciements à Monsieur Sébastien-Yves Laurent, mon directeur de thèse, qui a accepté de diriger cette thèse peu conventionnelle pour sa confiance, sa disponibilité et son encadrement. Je remercie également Monsieur le Lieutenant-Colonel Jérôme Barlatier et Monsieur Christophe Fagot, qui ont supervisé mon travail respectivement à la Gendarmerie et dans mon entreprise d'accueil, et dont la bienveillance, l'expertise et les remarques ont été précieuses. Je leur dois l'accès à des données de terrain et la signature d'un contrat de thèse CIFRE ayant financé ces recherches.

J'adresse ma gratitude à Messieurs Y. Jounot et Didier Bazalgette, sans qui ce projet ne se serait pas concrétisé. Je remercie également les membres du jury, Messieurs Jérôme Barlatier, Xavier Crettiez, Christophe Fagot, Sébastien-Yves Laurent, Guy Mélançon, Bruno Pinaud, Olivier Ribaux et Gilles Venturini d'avoir accepté de passer cette thèse au crible de leur expertise, ainsi que Gautier Cloix d'avoir accepté de siéger au jury en tant que membre invité.

Je remercie de plus Monsieur Bruno Pinaud pour son intérêt face aux problématiques techniques que j'ai rencontrées, pour tout le travail conjoint effectué, son accompagnement et son apport de connaissances.

Mes remerciements vont également aux membres de mon comité de suivi, Madame Stéphanie Brossard, Messieurs Bruno Pinaud, Jean-Marc Cochet, Boris Guarisma et le Général Demezou, à Carole Bergerot et aux membres de l'Institut de Recherche Montesquieu, ainsi qu'aux membres de l'école doctorale Montaigne Humanités pour leur temps, leur écoute et leurs conseils.

Merci à Benjamin et à Marcel pour leurs encouragements, leur soutien sans faille ces six dernières années et leurs innombrables relectures et corrections. Merci d'avoir été là dans les moments de doute et de m'avoir aidée à prendre du recul sur mon travail quand c'était nécessaire.

Merci à Nicolas Merli et Arnaud Tournier pour les heures passées à discuter de mon sujet et toutes les idées qui ont émergé de ces échanges. Merci à Laurence Benichou pour la rigueur de ses relectures et sa patience.

Merci à ma famille, Michel, Anne, Caroline et Nicole pour leur soutien indéfectible, et pour m'avoir permis d'en arriver là.

Merci enfin à mes amis, Léo, Laurie, Sylvain, Charles, Victor, Pierre, Céline – et tous ceux qui ne sont pas cités mais qui se reconnaîtront – pour leur présence, leur optimisme, leur bonne humeur, leurs encouragements ainsi que le soutien et la force qu'ils m'ont donnés.

À Michel, mon père qui m'a tant appris.

« L'art de la sécurité, c'est l'ensemble des moyens méthodologiques, des pratiques et des savoirs élaborés par des spécialistes des secteurs publics et privés ayant reçu pour mission de protéger et sécuriser les personnes, les biens et les institutions. »¹

Maurice Cusson

¹ Maurice Cusson, *L'art de la sécurité : ce que l'histoire de la criminologie nous enseigne*, Lausanne, Presses polytechniques et universitaires Romandes, 2011, 248 p., p. 11.

INTRODUCTION

Le renseignement criminel est une approche proactive de la fonction de police appuyant l'ensemble des forces de sécurité intérieure. Par opposition au caractère réactif et la visée à court terme de l'enquête criminelle, cette approche s'inscrit dans le qualitatif et le moyen terme². Il a pour objet l'étude des phénomènes criminels tels qu'ils se produisent afin d'en comprendre les mécanismes, et ce, à des fins d'anticipation et de mise en place de solutions proactives visant à endiguer ces phénomènes³. Il est essentiel de connaître la nature d'un mal pour le combattre efficacement⁴.

Si cette idée apparaît aujourd'hui comme une évidence, cela n'a pas toujours été le cas dans la lutte contre la délinquance et la criminalité, comme le retrace l'histoire finalement récente de la police telle que nous la connaissons, c'est-à-dire depuis sa séparation de la Justice.⁵

Afin de bien comprendre comment se positionne le renseignement criminel parmi les différents modèles de police, nous proposons en introduction de cette thèse de doctorat d'en exposer les prémisses.

Une histoire de la police guidée par le renseignement : les exemples de la France, du Royaume-Uni et des États-Unis.

Au XVII^{ème} siècle, sous le règne de Louis XIV, la France connaît une urbanisation massive, avec notamment une montée de la population à Paris. L'augmentation de la population parisienne entraîne dans la capitale une hausse des opportunités criminelles qui s'accompagne rapidement de troubles de l'ordre et de la sécurité : vols, incendies, menaces armées... Le roi, conscient que ces désordres engendrent de plus en plus d'insécurité à Paris, est soucieux pour sa couronne et sa sécurité.

Dans ce contexte, Jean-Baptiste Colbert, alors contrôleur général des Finances (poste qu'il occupe entre 1664 et 1683), envisage une approche globale de la criminalité et propose au roi un édit royal qui constituera l'acte fondateur de la police sous l'Ancien Régime. En 1667,

² Jean-François Gayraud et François Farcy, *Le renseignement criminel*, Paris, CNRS éd, 2014, 136 p.

³ Maurice Cusson, Frédéric Lemieux et Benoît Dupont, *Traité de sécurité intérieure*, Lausanne, Presses polytechniques et universitaires romandes, 2007, 705 p.

⁴ Jean-François Gayraud et François Farcy, *Le renseignement criminel*, *op. cit.*

⁵ Jean-Marc Berlière et René Lévy, *Histoire des polices en France : de l'Ancien régime à nos jours*, Paris, Nouveau monde éd, 2013.

Louis XIV crée par édit royal la première police de France, cristallisant ainsi l'idée d'un corps professionnel et d'une institution menant une action de sécurité liée à l'État⁶ (figure 1). Le magistrat Nicolas de La Reynie est nommé lieutenant général de police par le roi, poste qu'il occupera jusqu'en 1697. Doté de pouvoirs quasiment illimités, il a pour tâche de « purger la ville de ce qui peut causer des désordres »⁷ avant que la situation ne se dégrade et ne puisse menacer la couronne : propreté de la ville, contrôle du port d'armes, gestion des incendies et inondations, approvisionnement de la cité, activité des corporations⁸...

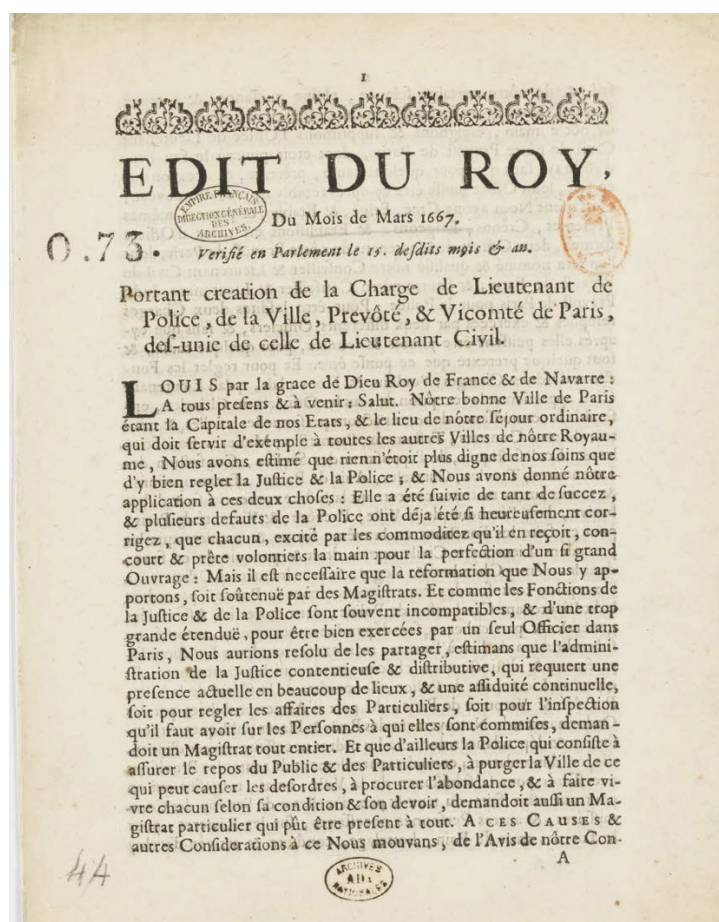


Figure 1 Édit de Louis XIV portant création de la lieutenance générale de police⁹

Dans un premier temps, La Reynie va rendre les rues plus sûres : sa police surveille, patrouille, anticipe, prévient et intervient avant que l'incident ne dégénère. Elle est au plus proche de là où les désordres se produisent¹⁰. Certains dispositifs sécuritaires sont mis en place : changement des enseignes des commerçants afin qu'elles ne chutent plus sur les passants,

⁶ *Ibid.*

⁷ Édit imprimé portant création de la lieutenance générale de police, édité en mars 1667 par Louis XIV.

⁸ Isabelle Bernier, *Histoire: la police à Paris sous l'Ancien Régime*, <https://www.futura-sciences.com/sciences/questions-reponses/epoque-moderne-histoire-police-paris-sous-ancien-regime-12299/>, consulté le 30 mars 2022.

⁹ Édit de mars 1667 : création de la lieutenance générale de police de Paris par Louis XIV. Archives nationales de France, cote AD 390. © Wikimedia Commons, domaine public.

¹⁰ Maurice Cusson, Frédéric Lemieux et Benoît Dupont, *Traité de sécurité intérieure*, *op. cit.*

éclairage des rues¹¹... Mais, bien que fondamentalement proactive, la police de La Reynie est guidée par l'idée de protéger le trône du roi plus que par la sécurité de la population, dont la protection n'est pas en réalité une fin en soi, mais un moyen de maîtriser les troubles et désordres. Ainsi, le lieutenant général de la police va plus loin et met en place un dispositif de renseignement fonctionnant sur la base d'informateurs : les « mouches ». Ces informateurs, généralement anonymes, sont des observateurs dans les rues, les lieux de rencontres et les commerces : ils rapportent le quotidien et l'actualité de lieux opaques pour la police malgré leur caractère public. Ils reportent ce qui s'y passe, ce qui s'y dit, ce qui s'y pense : il est attendu d'eux qu'ils relatent non seulement les bruits et nouvelles qui circulent, mais aussi les propos, discussions et jugements qu'ils entendent¹². L'absolutisme de Louis XIV déteint sur sa police¹³ : ses « mouches » sont des gens du peuple et le roi en use et abuse, notamment en exagérant leur présence auprès de la population afin d'imposer une crainte du dispositif policier ; sa police devient une police de l'opinion cherchant à réprimer les risques pour le pouvoir (écrivains, imprimeurs, protestants : les parisiens ne disent plus, n'écrivent plus, ne publient plus ce qu'ils veulent), sa police est omniprésente, omnipotente, corrompue, ingérente et intrusive et met en danger les libertés individuelles. L'instrumentalisation de la police pour protéger l'État crée un sentiment d'insécurité dans la capitale.

Face à la répression des libertés et au climat de peur et de méfiance instaurés en France, les Anglais se refusent à voir apparaître sur leur territoire une police d'État : « Les Anglais étaient fiers de leur absence d'officiers de police car la police était associée aux Français, et ainsi à la tyrannie¹⁴ ». De manière générale, les monarques et ministres, autant que le peuple, sont convaincus des vertus des « libertés à l'anglaise », perçues comme héritage de la *Bloodless Revolution* de 1688¹⁵ : malgré l'insécurité perçue, peu de décideurs sont enclins à entreprendre une réorganisation du dispositif de sécurité urbaine¹⁶. Ainsi, au XVIIIème siècle, le système judiciaire anglais est en crise et débordé. La Justice, dans un but dissuasif, doit faire peur : les

¹¹ Jean-Marc Berlière et René Lévy, *Histoire des polices en France*, op. cit.

¹² Gilles Malandain, « Les mouches de la police et le vol des mots : Les gazetins de la police secrète et la surveillance de l'expression publique à Paris au deuxième quart du XVIIIe siècle », *Revue d'histoire moderne et contemporaine*, vol. 42, n° 3, 09/1995, pp. 376-404.

¹³ Olivier Ribaux, cours « Policing », ESC – Université de Lausanne, 2015.

¹⁴ Thomas Keneally, *The commonwealth of thieves: the Sydney experiment*, Milsons Point, NSW, Random House Australia, 2005, 509 p., p31. Citation originale : “*The English were proud of their lack of police officers because police was associated with the French, and thus with tyranny*” traduite de l'anglais vers le français par l'auteure.

¹⁵ The *Bloodless Revolution*, ou *Glorious Revolution*, restaura une dynastie protestante sur le trône d'Angleterre, d'Irlande et d'Écosse lorsque le Roi James II d'Angleterre, catholique, fut destitué et que le Parlement appela sur le trône sa fille, Marie II et son époux William III, prince d'Orange, tous deux protestants. Cela eut pour conséquence de renforcer la monarchie parlementaire et le rôle du Parlement face à la Couronne.

¹⁶ Clive Emsley et Pierre Gervais, « Police, maintien de l'ordre et espaces urbains : une lecture anglaise », *Revue d'histoire moderne et contemporaine*, vol. 50, n° 1, 2003, pp. 5-12.

peines sont volontairement excessives pour de nombreux délits, les prisons – pleines – sont dans un état désastreux, et à la fin du siècle, suite à la guerre d’indépendance d’Amérique du Nord, il n’est plus possible d’envoyer les « *convicts* »¹⁷ outre-Atlantique. À partir de 1788, certains d’entre eux sont ainsi déportés en Australie, toute juste découverte.

Pour faire face aux tensions du système, une professionnalisation du dispositif de sécurité apparaît progressivement : création des « *constables* » et des « *watchmen* »¹⁸. Les premiers étaient élus pour un an, non rémunérés, et avaient notamment pour charge le prélèvement des taxes, les vérifications des poids et mesures et les arrestations ordonnées par le gouvernement ; tandis que les seconds étaient désignés parmi les hommes physiquement capables ayant entre 25 et 55 ans, et étaient chargés des patrouilles et du maintien de l’ordre¹⁹. Avec ce dispositif bancal, lacunaire et inégal, l’enquête policière n’existe pas encore : il revient aux victimes d’infractions elles-mêmes d’investiguer et de présenter le coupable à la justice, à moins qu’elles n’aient les moyens de rémunérer le *constable* pour une prise en charge prioritaire ou de faire appel à des organisations privées (*thief takers, bow street runners*). En outre, la posture anglaise à l’égard des droits et libertés individuels contraint considérablement les possibilités de supervision et de contrôle de l’espace public²⁰.

À cette période, l’Angleterre vit beaucoup du commerce qui transite par le port de la Tamise, constituant un goulot d’étranglement dans lequel sont concentrées énormément de marchandises et de richesses, et représentant une importante opportunité criminelle en termes de vols et détournements de marchandises²¹.

Ainsi, en 1798, Patrick Colquhoun²², marchand et statisticien, met en place la « Police de la Tamise », une police privée financée par les marchands dont les cargaisons transitent par le port. Il s’agit du premier mouvement de ce qui deviendra la toute première police de prévention. En effet, partant du principe qu’il n’est pas efficace de lutter contre un phénomène qui n’est pas compris, Colquhoun travaille sur les types de problèmes, les modes opératoires, les situations (contraintes et circonstances immédiates aux vols), les types d’auteurs

¹⁷ Les « *convicts* » sont les individus emprisonnés ou déportés par la justice anglaise au XVIIIème et XIXème siècles.

¹⁸ Philippe Chassaigne, « Criminalité et police à Londres au XIXe siècle », *Revue française de civilisation britannique*, vol. 12, n° 3, 2003, pp. 2-7.

¹⁹ Maurice Cusson, *L’art de la sécurité*, *op. cit.*

²⁰ Clive Emsley et Pierre Gervais, « Police, maintien de l’ordre et espaces urbains : une lecture anglaise », *op. cit.*

²¹ Leon Radzinowicz, *A history of English criminal law and its administration from 1750*, London, Stevens, 1948.

²² Patrick Colquhoun est considéré comme le fondateur de la première force de police régulière d’Angleterre. Marchand et statisticien écossais, il fonde la Chambre de commerce de Glasgow en 1783 avant de partir s’installer à Londres pour poursuivre sa carrière, où il deviendra magistrat de police.

(travailleurs des docks, clients, marchands, voleurs...) et veut connaître les méthodes de manière générale²³.

Rapidement, grâce à son travail d'analyse, il identifie des *patterns* dans les opportunités ainsi que des modes opératoires spécifiques, dont il va utiliser les caractéristiques pour dégager des solutions proactives taillées sur mesure²⁴. Un *pattern* est un motif, ou modèle simplifié, d'une structure qui se répète. En renseignement criminel sont distingués les *patterns* d'activité, qui sont des « répétitions criminelles qui présentent une structure, une trame dans l'activité²⁵ », des *patterns* dans les données, qui sont « des structures repérées dans les données collectées²⁶ ». Les premiers sont des régularités existantes dans la criminalité, tandis que les seconds en sont les conséquences visibles dans les données.

Ainsi, l'analyse du problème conduit vers la solution : les bases des modèles actuels de police guidée par le renseignement (*ILP – Intelligence-Led Policing*) sont posées²⁷. Colquhoun rédige alors deux traités²⁸ sur la police, dans lesquels il développe la notion de « science de la police » à une époque où la police n'existe pas encore réellement. Ses traités inspirent notamment la fondation de la police à Sydney, Dublin et New York²⁹.

Le modèle est bon, fonctionne, et est apprécié de ses clients – essentiellement des commerçants de la Compagnie des Indes occidentales – qui jugent le succès de la police de la Tamise immédiat³⁰. Deux ans après sa création, en 1800, cette activité devient publique. La police de Londres, la *Metropolitan Police*, est créée par Robert Peel³¹ en 1829 et intègre le modèle de Colquhoun en 1839.

La création de la *metropolitan* s'accompagne de la rédaction par le Parlement du *Metropolitan Police Act of 1829* (ou *Metropolitan Act*), un texte qui entend répondre aux préoccupations de la population d'une part concernant les désordres et la criminalité

²³ Patrick Colquhoun (1800), *A Treatise on the Commerce and the Police of the River Thames*, Montclair, Patterson Smith, 1969, 712 p.

²⁴ Olivier Ribaux, cours « *Policing* », *op. cit.*

²⁵ Olivier Ribaux et Quentin Rossy, cours « Analyse criminelle opérationnelle », ESC – Université de Lausanne, 2015.

²⁶ *Ibid.*

²⁷ *Ibid.*

²⁸ Patrick Colquhoun (1800), *A Treatise on the Commerce and the Police of the River Thames*, *op. cit.* ; Patrick Colquhoun (1806), *Treatise on the Police of the Metropolis*, Montclair, Patterson Smith, 1969, 519 p.

²⁹ Olivier Ribaux, cours « *Policing* », *op. cit.*

³⁰ Leon Radzinowicz, *A history of English criminal law and its administration from 1750*, *op. cit.*

³¹ Robert Peel est un homme d'État britannique du XIX^{ème} siècle. Fondateur de la *metropolitan* police de Londres, il est secrétaire d'État à l'Intérieur de 1822 à 1827, puis premier ministre du Royaume-Uni de 1834 à 1835 et de 1841 à 1846.

grandissants, d'autre part de voir apparaître une police basée sur le modèle français. Le *Metropolitan Act* fonde le Bureau du commissaire de police (*office of the Commissioner of Police*), en charge de la gestion et de l'administration du corps policier. Ce texte met en place des bases fondamentalement différentes du modèle français : professionnalisation des forces de police, redevabilité et droit de regard du public sur ses agissements, établissement de commissions d'enquête en cas de mauvaise conduite des policiers, désarmement des effectifs de patrouille. Une liste de neuf principes, souvent attribués à Robert Peel et désignés sous le nom « *Peel's nine police principles* »³² rassemble les concepts fondamentaux de la *metropolitan* : une police communautaire, partie intégrante de la population et agissant en coopération avec cette dernière (cf annexe 2). De ces principes découle la construction du modèle de la police anglaise comme police idéale, dont l'image n'est plus celle de l'exercice d'un pouvoir arbitraire, mais d'une légalité neutre et d'une police proche de la situation et de ses problèmes. Ce fonctionnement se trouve renforcé par la prolifération des contraintes légales sur les procédures que la police pouvait et devait employer pour maintenir l'ordre et faire appliquer les lois³³.

Cependant, après plus de cent ans de bon fonctionnement, ce modèle finit par s'enliser et entre en crise à la fin du XXème siècle : les crimes et délits déclarés à la police augmentent bien plus rapidement que les effectifs, et les procédures administratives prolifèrent³⁴. Le temps manque pour la prévention et la proactivité et le dispositif devenu réactif est de moins en moins efficace. En 1993, un audit du système policier conclut à un « cercle vicieux d'échecs dans la lutte contre le crime³⁵ », et la police est réformée en profondeur : nettoyage administratif, création de *crime desks* pour mieux gérer les interventions (évaluations de situation, suivi d'investigations et maintien du contact avec les victimes)³⁶, lutte contre les crimes répétitifs, ciblage des auteurs prolifiques, étude de la criminalité...

³² Bien que ces neuf principes soient très souvent dans la littérature attribués à Robert Peel, une étude réalisée en 2007 par Susan A. Lentz et Robert H. Chaires (Susan A. Lentz et Robert H. Chaires, « The invention of Peel's principles: A study of policing 'textbook' history », *Journal of Criminal Justice*, vol. 35, n° 1, 2007, pp. 69-79) montre qu'aucun document ou source primaire ne permet effectivement de les relier à Robert Peel. Selon les auteurs, ces neuf principes auraient été formalisés plus tardivement, dans le courant du XXème siècle, et seraient en réalité un « résumé » de l'évolution au XIXème siècle de la police en Angleterre.

³³ Robert Rainer, « La tradition policière britannique : modèle ou mythe ? », *Cahiers de la sécurité intérieure*, n° 7, 1992/1991, pp. 27-39.

³⁴ Une arrestation pour vol entraîne un total de 40 formulaires à remplir.

³⁵ Audit Commission for Local Authorities and the National Health Service in England and Wales, *Helping with enquiries: tackling crime effectively*, London, HMSO, 1993, 65 p.

³⁶ *Ibid.*

Sont notamment reprises en Angleterre des études menées aux États-Unis depuis les années 1970 sur des cohortes de jeunes hommes issus de différentes régions nés en 1953³⁷ ainsi qu'en 1945 et 1958³⁸ et suivis « juridiquement » pendant plusieurs années. Les résultats sont similaires à ceux des études américaines : 33% des hommes de la cohorte ont été condamnés, dont 20% (soit 7% de la cohorte) pour 65% des infractions jugées³⁹. La délinquance est répétitive, et la police procède à des investigations mieux gérées, ciblant les auteurs prolifiques et s'intéressant à leurs trajectoires criminelles⁴⁰. Outre les auteurs prolifiques, deux autres concentrations sont constatées : d'une part, une étude effectuée à Minneapolis sur les « points chauds » montre que, en un an, sur les 233 979 appels à la police de la cohorte étudiée, 50% sont concentrés dans 3% de la zone géographique couverte (115 000 localisations), et que tous les cas de vols aggravés, de viols et de vols de voitures sont respectivement concentrés dans 2,2%, 1,2% et 2,7% de la zone géographique⁴¹. D'autre part, des enquêtes sur les victimes mettent en évidence que les mêmes victimes (ou types de victimes) subissent les crimes : plus la victimisation est répétée, plus la probabilité d'être à nouveau une cible est élevée. La police retrouve sa volonté préventive et proactive et concentre ses efforts sur ces « points chauds », sur la détermination de profils d'auteurs prolifiques, la compréhension des typologies de victimes ciblées et les moyens de protection à mettre en place⁴². Le développement du *crime pattern analysis* donne lieu à une construction progressive du système d'*Intelligence-Led Policing (ILP)*, qui atteint son point culminant avec la création du *National Intelligence Model (NIM)* en 1999. Ce modèle est mis en place par le *National Criminal Intelligence Service (NCIS)*⁴³, à la suite du constat de la commission d'audit qu'il existe une forte disparité nationale dans les pratiques de renseignement, entravant considérablement le flux d'informations à tous les niveaux (local, régional et national)⁴⁴.

³⁷ Marvin E. Wolfgang, Robert M. Figlio et Thorsten Sellin, *Delinquency in a Birth Cohort*, University of Chicago Press, 1972, 116 p.

³⁸ Marvin E. Wolfgang et Paul E. Tracy, « The 1945 and 1958 birth cohorts: a comparison of the prevalence, incidence, and severity of delinquent behavior », Philadelphia, Pa., Center for Studies in Criminology and Criminal Law, University of Pennsylvania, 1982.

³⁹ Ministry of Justice, *Conviction histories of Offenders between the ages of 10 and 52*, London, Ministry of Justice, 2010, 16 p.

⁴⁰ Olivier Ribaux, cours « Policing », *op. cit.*

⁴¹ Lawrence W. Sherman, Patrick R. Gartin et Michael E. Buerger, « Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place », *Criminology*, vol. 27, n° 1, 02/1989, pp. 27-56.

⁴² Ronald V Clarke et John E Eck, *Crime Analysis for Problem Solvers in 60 Small Steps*, Washington, DC, U.S. Department of Justice, COPS, 2005, 150 p.

⁴³ Le NCIS, crée en avril 1992, avait pour fonction de recueillir des données et de les analyser afin de fournir aux forces nationales les informations et renseignement dont elles avaient besoin. Le 1^{er} avril 2006, ce service a été fusionné avec le *National Crime Squad (NCS)*, une agence qui traitait les crimes organisés nationaux et transnationaux, pour devenir la *Serious Organised Crime Agency (SOCA)*.

⁴⁴ Audit Commission for Local Authorities and the National Health Service in England and Wales, *Helping with enquiries*, *op. cit.* ; Association of Chief Police Officers, *Guidance on the National Intelligence Model*, National Centre for Policing Excellence, 2005, 213 p.

Le *NIM* puise ses racines dans le renseignement criminel et se fonde sur « la sagesse collective et les meilleures pratiques »⁴⁵ ; il se dote en outre d'un modèle de processus opérationnels visant à faciliter l'organisation des connaissances et la gestion des ressources, à informer des menaces, de leurs gravités et niveaux de risques, à coordonner les activités sur l'ensemble des questions policières, du crime organisé à la sécurité routière et à réviser les tactiques policières⁴⁶.

L'application de ce modèle réduit les obstacles qui s'opposent à l'efficacité du système en produisant des processus et un langage normalisés pour créer un environnement de travail coopératif. Le renseignement est au cœur de la planification des activités, qui tient compte des objectifs locaux et gouvernementaux et des principes d'optimisation des ressources. L'élément clef d'une planification réussie est la compréhension de cinq aspects : ce qui se passe réellement sur le terrain, la nature et l'étendue des problèmes, les tendances et la cartographie des principales menaces. Le processus opérationnel, également fondé sur le renseignement, concerne le déploiement proactif de ressources pour réduire les activités frauduleuses et les crimes, ainsi que les problèmes associés qui nuisent à la sûreté et la sécurité.

Aux États-Unis, cette même idée de l'importance de comprendre la délinquance et la criminalité naît à Berkeley en 1909 avec la nomination d'August Völlmer en tant que chef de la police. Malgré un court parcours scolaire, Völlmer a soif de connaissances : autodidacte, il s'instruit en allant à la rencontre de professeurs et en dépouillant les ouvrages de l'Université de Californie, à Berkeley. Il deviendra une telle figure de référence en matière de connaissances et d'éducation policières qu'il publiera des articles dans des revues liées au maintien de l'ordre et sera invité dès 1929 à enseigner à l'Université de Chicago en tant que professeur. Considéré comme un précurseur avant-gardiste de la police, son influence se généralisera d'abord à l'ensemble de la Californie puis des États-Unis avant d'atteindre les autres continents.

Secondé par Orlando Winfield Wilson, diplômé de l'Université de Californie en criminologie, Völlmer bouleverse fondamentalement la police en apportant beaucoup d'idées mais aussi en recrutant dans la police de Berkeley des universitaires, surnommés les « *college cops* ». Dans trois ouvrages⁴⁷, Wilson formalise ces idées qui affectent tous les aspects de la

⁴⁵ Association of Chief Police Officers, *Guidance on the National Intelligence Model*, *op. cit.*

⁴⁶ *Ibid.*

⁴⁷ O.W. Wilson, *Police Records Their Installation and Use*, Berkeley, Public Administration Service, 1951, 376 p. ; O. W. Wilson, *Police planning*, Springfield, Ill., Thomas, 1977 ; O. W. Wilson et Roy Clinton McLaren, *Police administration*, New York, McGraw-Hill, 1972, 742 p.

police, avec notamment la création du premier laboratoire de police et de police scientifique⁴⁸. Sous la supervision de Völlmer, les « *college cops* » mettent en place un fichier d'empreintes digitales initialement alimenté par les empreintes de suspects et criminels. En 1936, le Département de Police de Berkeley réalise une campagne de « *fingerprinting* », encourageant les citoyens à venir donner leurs empreintes digitales, et collecte cinquante mille empreintes⁴⁹. Völlmer professionnalise la police et met en place des standards de recrutement et de formation, allège les processus administratifs, accroît l'efficacité des opérations, procède à une répartition claire des responsabilités ainsi qu'à une meilleure utilisation des ressources humaines, et dote la police des moyens technologiques de son temps⁵⁰ en motorisant les patrouilles, en instaurant l'utilisation de la radio pour les communications ou en initiant l'invention du « polygraphe⁵¹ ».

*« The time has come when the official police department may afford frankly to concede that Science can help them to apprehend the criminal. »*⁵²

August Völlmer

L'idée d'un traitement rationnel des problèmes se développe : constatation et analyse des concentrations du crime, mise en place d'un système de gestion géographique en punaisant sur une carte les lieux d'où sont émis les appels à la police, orientation proactive des patrouilles sur la base de ces informations, étude des modes opératoires, campagnes de prévention sur les jeunes, contrôle centralisé... Les actions policières reflètent non seulement une analyse, mais aussi une grande rationalité sur les problèmes.

De par sa conception globale, Völlmer incite tout au long de sa carrière ses collègues officiers de police à publier les méthodes et outils qu'ils mettent en œuvre⁵³. Ses réformes imposent un tel changement et induisent une telle efficacité qu'Holstrom⁵⁴ dira que Berkeley

⁴⁸ Olivier Ribaux, *Police scientifique : le renseignement par la trace*, Lausanne, Presses polytechniques et universitaires romandes, 2014.

⁴⁹ Simon A. Cole, *Suspect identities: a history of fingerprinting and criminal identification*, Cambridge, MA, Harvard Univ. Press, 2002, 369 p.

⁵⁰ Olivier Ribaux, *Police scientifique*, *op. cit.*

⁵¹ Ce « détecteur de mensonges » est inventé par John Larson, doctorant à l'Université de Californie, à la demande de Völlmer et mis en service dès 1929.

⁵² Regional Oral History Office, *August Vollmer: Pioneer in Police Professionalism: Oral History Transcript/ and Related Material*, Berkeley, University of California, 1972.

⁵³ August Vollmer, « Revision of the Atcherley Modus Operandi System », *Journal of the American Institute of Criminal Law and Criminology*, vol. 10, n° 2, 1919, pp. 229-274 ; August Vollmer, « The Scientific Policeman », *The American Journal of Police Science*, vol. 1, n° 1, 01/1930, pp. 8 ; O. W. Wilson, « August Vollmer », *The Journal of Criminal Law, Criminology, and Police Science*, vol. 44, n° 1, 05/1953, p. 91 ; Regional Oral History Office, *August Vollmer*, *op. cit.*

⁵⁴ John Holstrom a été chef de la police de Berkeley de 1944 à 1960.

est passée de la ville la moins sûre des États-Unis avant l'arrivée de Völlmer à la direction de la police à la plus sûre à son départ⁵⁵.

Dans la continuité du rapprochement entre policiers et monde académique, il s'associe avec Paul L. Kirk⁵⁶ afin d'initier la création d'une École de Criminologie à l'Université de Californie⁵⁷. Le projet se concrétise et l'École est fondée en 1950⁵⁸ autour de trois piliers fondateurs : la criminalistique (forensique), le système correctionnel et l'action de sécurité⁵⁹. Dix ans plus tard, en 1960, l'École de Criminologie de l'Université de Montréal voit le jour.

Aujourd'hui, la capitalisation de connaissances sur les phénomènes criminels, les *modus operandi*, les auteurs, les victimes, les points chauds, *etc.*, apparaît complémentaire et indissociable de l'enquête criminelle pour assurer les missions de sécurité et de sûreté des forces de police.

Historique de la recherche : l'élaboration du sujet.

Le présent travail de modélisation de réseaux criminels s'inscrit dans le cadre de l'ILP et du renseignement criminel, avec de profondes convictions dans le potentiel des recherches académiques en criminologie.

Lors de mon apprentissage en Master en Sciences ès Sciences criminelles à l'École des Sciences criminelles de l'Université de Lausanne de 2014 à 2017, de nombreuses compétences et connaissances en police scientifique ont été acquises et développées : effectuer des prélèvements de traces papillaires et de traces de semelles, effectuer des analyses ADN, rechercher l'arme source de douilles et projectiles tirés, déterminer le profil chimique de produits stupéfiants, caractériser un départ d'incendie, interpréter les indices scientifiques, évaluer les preuves et leurs pertinences... Nombreuses sont les causes qui nécessitent l'intervention de la Police Technique et Scientifique (PTS) afin de retracer des événements, des

⁵⁵ John Holstrom, « Vollmer as a man: Memories of a Close Friend and Colleague », dans *August Vollmer: Pioneer in Police Professionalism: Oral History Transcript/ and Related Material*, R.O.H. Office., Berkeley, University of California, 1972.

⁵⁶ Biochimiste et criminaliste américain, Kirk a notamment pu fournir les preuves essentielles des éclaboussures de sang qui ont conduit à l'acquittement de Sam Sheppard, incarcéré pour le meurtre de sa femme, lors d'un procès plus de 12 ans après les faits. La plus haute distinction que l'on puisse recevoir dans la section criminalistique de l'*American Academy of Forensic Sciences* porte son nom.

⁵⁷ John Holstrom, « Vollmer as a man: Memories of a Close Friend and Colleague », *op. cit.*

⁵⁸ Malgré le succès de ses filières

⁵⁹ Paul Leland Kirk, « Standardization of Criminological Nomenclature », *Journal of Criminal Law & Criminology*, vol. 38, n° 2, 1947, pp. 165-167.

actes, et d'en identifier les auteurs.

Au cours de ce cursus, parmi toutes les disciplines qui soutiennent l'enquête, une approche s'est distinguée : l'approche proactive que proposent les enseignements d'« analyse criminelle opérationnelle (ACO) » et de « renseignement forensique » – par opposition à la démarche intrinsèquement réactive de l'enquête et des investigations judiciaires. L'objet de ces deux disciplines est d'étudier la criminalité telle qu'elle se manifeste, afin d'atteindre un meilleur niveau de compréhension de l'existant : il ne s'agit plus d'identifier les individus auteurs de faits répréhensibles dans le but de les soumettre au système judiciaire, mais de s'intéresser aux phénomènes criminels en tant que tels, de s'attacher à en comprendre les fonctionnements, forces et vulnérabilités afin d'adapter la réponse opérationnelle et stratégique.

*« L'enquête criminelle cherche les auteurs et les faits (qui, quoi),
le renseignement criminel cherche les mécanismes (comment). »*

Jérôme Barlatier

Les enseignements des modèles d'ILP anglais et américains montrent que, pour dégager et mettre en place des solutions proactives adaptées aux problèmes rencontrés, l'étude et l'analyse des types de problèmes, des typologies de victimes et d'auteurs, des situations, des modes opératoires et des stratégies criminelles s'avèrent efficace. En effet, cette approche permet de détecter des *patterns*, *i.e.* des schémas qui se répètent dans les données et qui illustrent des récurrences dans les opportunités, les activités et les méthodes criminelles.

*« Il est d'une importance extrême dans la guerre d'attaquer la
stratégie de l'ennemi. »⁶⁰*

Sun Tzu

Sur la base des *patterns* dans les données ainsi détectés, il est essentiel de dégager des *patterns* d'activités, *i.e.* de comprendre comment les stratégies sont mises en place, comment et pourquoi elles fonctionnent et comment elles permettent aux criminels de perpétuer leurs activités et ce, malgré les systèmes de lutte contre ces phénomènes qui peuvent être mis en place. Seule une connaissance des mécanismes criminels mis en œuvre permettra d'en déterminer les forces et les vulnérabilités, et ainsi de lutter efficacement contre. À travers

⁶⁰ Sun Tzu, *L'art de la guerre : Mille et une nuits.*, Paris, Pocket, 2000, 128 p., p. 5.

l'approche renseignement criminel, la lutte contre le crime est pensée sous l'angle de la réflexion stratégique et de l'anticipation⁶¹.

Dans cette optique, d'étude et de compréhension de mécanismes, ce travail n'aurait pu voir le jour sans s'appuyer sur des données concrètes. En effet, bien qu'un grand nombre d'études théoriques et bibliographiques d'une grande qualité aient beaucoup contribué à la recherche et à faire avancer l'état des connaissances, cette thèse se veut empirique. Plus encore, l'un de ses objectifs est d'apporter des éléments de méthode et d'analyse applicables par les forces de sécurité intérieure. Ainsi, cette recherche consiste en l'étude d'un cas d'usage pratique et réel : une enquête de gendarmerie sur un réseau de trafic de fonds et de blanchiment ayant agi sur l'ensemble du territoire français jusqu'à son démantèlement en 2018. Le matériau de travail est constitué de l'ensemble des pièces de procédures de l'enquête ainsi que des matrices d'analyse de données confectionnées par les analystes ayant pris part aux investigations. Pour des raisons de confidentialité, bien que cette affaire ait aujourd'hui – et depuis le début des travaux en juin 2019 – « autorité de chose jugée », toutes les informations à caractère personnel seront anonymisées avant toute diffusion de ces travaux, y compris le présent manuscrit.

Cette thèse de doctorat, « Modélisation hybride de réseaux dans un « champ criminel » : contribution des sciences sociales et d'outils logiciels au renseignement criminel », entend mettre l'accent sur la compréhension des mécanismes sociaux et du fonctionnement de l'organisation criminelle étudiée sur la base du cas d'usage. Le couplage des sciences sociales et d'outils logiciels introduit le caractère hybride de cette recherche. En effet, l'objectif, loin de se cantonner à une seule discipline, est d'utiliser les outils analytiques que sont notamment les statistiques, les mathématiques et l'informatique pour transformer la donnée en information exploitable par les sciences humaines et sociales. La construction du sujet a été effectuée conjointement avec les parties prenantes de ce projet.

Lors de sa mise en place, chacune des parties a apporté une approche intéressante, venant enrichir la pluridisciplinarité déjà ambitionnée de la recherche⁶². En effet, ce projet de doctorat débute en 2017 à la suite d'un parcours universitaire pluridisciplinaire : d'abord, une Licence en Sciences et Technologies, mention Physique & Interfaces, obtenue en 2014 à l'Université Pierre et Marie Curie (UPMC – Paris VI), puis un Master en Sciences ès Sciences criminelles à l'École des Sciences criminelles de l'Université de Lausanne obtenu en 2017.

⁶¹ Jean-François Gayraud et François Farcy, *Le renseignement criminel*, *op. cit.*

⁶² Cf. chapitre I « genèse du projet et contextualisation du sujet. »

L'apprentissage et la découverte de nouvelles disciplines lors du master a été une véritable ouverture d'esprit à de nouvelles méthodes, une nouvelle rigueur et de nouveaux modes de pensée. Ce parcours, certes atypique, tire sa force de la combinaison de méthodes et de raisonnements issus de champs disciplinaires aussi éloignés l'un de l'autre que semblent l'être la physique statistique et le renseignement forensique. La pluridisciplinarité devient un atout lorsque les disciplines se nourrissent mutuellement en s'apportant outils, arguments, méthodologies, idées et réponses.

Dans ce contexte de pluridisciplinarité, l'objectif initial était de travailler sur la modélisation de réseaux illicites dans une démarche de renseignement criminel. Au début de la thèse, en juin 2019, l'intitulé du sujet était le suivant : « Appréhension des réseaux criminels par les connaissances sociologiques et l'expertise des enquêteurs spécialistes : assistance dans l'analyse et les enquêtes sur les réseaux criminels par la modélisation systémique des approches criminelles et sociologiques ». Ce titre, assez lourd, entendait expliciter cette pluridisciplinarité : « appréhension des réseaux criminels », « assistance dans l'analyse et les enquêtes sur les réseaux criminels » pour introduire l'aspect renseignement dans sa fonction d'appui à l'enquête par les connaissances générales qu'il dégage. Venaient ensuite « les connaissances sociologiques », « l'expertise des enquêteurs spécialistes », « la modélisation systémique » et « [l]es approches criminelles et sociologiques » pour exposer les principales disciplines auxquelles fait appel cette thèse. Cependant, ce titre figeait la recherche dans son état initial et laissait peu de place à l'évolution du sujet. Or, un travail de recherche évolue nécessairement avec le temps, la recherche bibliographique, les sources et matériaux à disposition et les contraintes (temporelles, techniques, matérielles) rencontrées. Ainsi, l'intitulé a été modifié et allégé afin de laisser plus d'espace à l'évolution naturelle du sujet : « Modélisation des réseaux criminels par les connaissances humaines et sociales et l'expertise des enquêteurs spécialistes ». Dans la continuité de l'adaptation du sujet à son contexte de recherche, deux composantes centrales ont rapidement émergé de la recherche bibliographique : le concept de « champ social criminel », application de la théorie des champs sociaux de Pierre Bourdieu⁶³ au contexte spécifique de la criminalité, et l'étude de la structure d'une entité sociale⁶⁴. Outre ces deux aspects fondamentaux, l'accès aux données, *i.e.* la spécification du

⁶³ Pierre Bourdieu, « Espace social et genèse des "classes" », *Actes de la recherche en sciences sociales*, vol. 52-53, , 1984, pp. 3-14.

⁶⁴ Stephen P. Borgatti, Martin G. Everett et Jeffrey C. Johnson, *Analyzing social networks*, Los Angeles, SAGE, 2018, 363 p ; Ronald S. Burt, *Structural holes: the social structure of competition*, Cambridge, Mass., Harvard Univ. Press, 1995, 313 p ; Lucia Cavallaro, Ovidiu Bagdasar, Pasquale De Meo, Giacomo Fiumara et Antonio Liotta, « Graph and Network Theory for the analysis of Criminal Networks », *arXiv:2103.02504 [physics]*, 2021, pp. 139-156 ; Alain Degenne et Michel Forsé, *Les réseaux sociaux : une analyse structurale en sociologie*, Paris, A. Colin, 1994, 288 p ; Alexis Ferrand, « La structure des systèmes de relation », *L'Année sociologique*, vol. 47,

cas d'usage étudié, a amené une nouvelle évolution de l'intitulé : « Approche structurelle et modélisation outillée par les sciences sociales des réseaux dans le « champ criminel » du trafic et blanchiment de fonds en France. » Enfin, en novembre 2021⁶⁵, l'avancement de la recherche et la détermination de la méthode de travail a permis d'affiner le titre pour aboutir à l'intitulé final : « Modélisation hybride de réseaux dans un « champ criminel » : contribution des sciences sociales et d'outils logiciels au renseignement criminel. »

Ainsi, à travers l'évolution de l'intitulé, la mise en œuvre et la progression du sujet dans son ensemble sont exposées : outre les avancées bibliographiques, la découverte du cas d'usage pratique et le choix des méthodes employées, ces différentes versions ont permis de construire la démarche intellectuelle à suivre en ciblant les questions de recherche : l'organisation criminelle étudiée peut-elle être assimilée à un « champ criminel » spécifique ? Quels attributs sociaux et structuraux sont à considérer pour le déterminer ? L'étude de la structure sociale permet-elle de déterminer des critères de définition d'une fonction « stratégique » pour un réseau criminel ? Ces critères sont-ils décelables dans des données collectées et saisies dans une démarche d'investigation judiciaire ?

Dans ce cadre et en s'appuyant sur le matériel à disposition, cette thèse de doctorat entend répondre à la problématique suivante :

Peut-on combiner les outils analytiques issus de diverses disciplines pour développer, sur la base de données telles que collectées et saisies lors des enquêtes de gendarmerie, une méthodologie de représentation, de modélisation et d'analyse de réseaux criminels aboutissant à une détection des fonctions et positions stratégiques dans le fonctionnement de l'organisation ?

Pour y répondre, le premier chapitre abordera en détail la genèse de ce projet de doctorat, au travers des différents partenariats qui lui ont donné vie et de son aspect multidisciplinaire. Dans un deuxième chapitre, l'état d'avancement des connaissances dans les différentes disciplines auxquelles fait appel ce travail sera exposé. Le cas d'usage étudié, les données à disposition et leur structure seront décrits dans un troisième chapitre. Les contributions que se propose d'apporter ce manuscrit seront exposées dans les quatre chapitres suivants : la méthode

n° 1, 1997, pp. 37-54 ; Brigitte Gay, Bernard Dousset et Radwen Wanassi, « Les indicateurs structurels d'un graphe : calculs, visualisation, interactivité », 2013 ; Per Hage et Frank Harary, *Structural models in anthropology*, Cambridge [Cambridgeshire] ; New York, Cambridge University Press, 1983, 201 p ; Pierre Mercklé, *Sociologie des réseaux sociaux*, Paris, Découverte, 2004, 121 p ; Robert K. Merton, « Social Structure and Anomie », *American Sociological Review*, vol. 3, n° 6, 1938, pp. 672-678 ; Stanley Wasserman et Katherine Faust, *Social Network Analysis Methods and Applications*, Cambridge, Cambridge University Press, 1994, 826 p.

⁶⁵ Début de la troisième et dernière année de thèse.

de travail employée pour préparer les données, les traiter, les analyser et en extraire le réseau social de l'organisation sera explicitée dans les quatrième et cinquième chapitres. Le sixième sera consacré à la discussion de la méthode appliquée ; le septième proposera une description et une analyse du réseau criminel, ainsi qu'une interprétation des résultats d'un point de vue analytique. Le huitième chapitre s'emploiera à interpréter et discuter les résultats d'un point de vue sociologique. Enfin, la conclusion proposera un résumé de la démarche et appellera à de nouvelles perspectives de recherche.

Sommaire

CHAPITRE I	Genèse du projet et contextualisation du sujet
CHAPITRE II	État d'avancement des connaissances : des sciences sociales à la modélisation
CHAPITRE III	Cas d'usage et données d'étude : présentation d'une enquête d'envergure et structuration des données
CHAPITRE IV	Du pré-traitement des données brutes au graphe multivarié
CHAPITRE V	Méthodologie générale : du graphe multivarié au réseau social
CHAPITRE VI	Discussion de la méthode d'extraction et détermination du seuil de filtrage
CHAPITRE VII	Analyse de réseaux sociaux : application au sociogramme <i>INR</i>
CHAPITRE VIII	Interprétations en termes d'espace et de champ criminels : le coup de tonnerre du 10 Novembre
	<i>Conclusion</i>

CHAPITRE I GENÈSE DU PROJET ET CONTEXTUALISATION DU SUJET

I.1 Historique de la recherche : du projet au doctorat

Ce projet de doctorat est une initiative personnelle, née en 2016 d'une prise de conscience de l'ampleur des conséquences de la criminalité organisée sur la société dans laquelle elle opère. Ses effets – volontaires ou non, conscients ou non, perçus ou non – sont subis directement par les auteurs, les victimes et leurs proches, mais aussi indirectement par l'ensemble de la structure sociale. Sur la base de convictions profondes dans la nécessité de mieux comprendre la criminalité pour mieux lutter contre – issues des enseignements du Master en cours⁶⁶ –, ce projet a germé à l'état d'idée un peu chimérique : celle d'effectuer une thèse de doctorat comme il n'en existe pas dans une discipline qui n'existe pas au sens académique : le renseignement criminel.

Les questions pratiques qui se posent dès le mois de juin 2016 et les obstacles qui s'annoncent transforment l'idée chimérique en projet complexe : dans quelle discipline académique s'inscrire ? Dans quel laboratoire ? Qui serait intéressé à encadrer de telles recherches ? Où et comment les financer ? Sur quelles données travailler ? Mais surtout : quel sujet proposer ? « Étudier la criminalité organisée pour mieux la comprendre » ce n'est pas un sujet de thèse, c'est une envie. À ce stade, le champ de recherche n'est volontairement pas plus précis : il y a tant de possibilités d'axes et de sujets de recherches dans ce passionnant domaine d'application qu'il est peut-être plus bénéfique de ne pas trop se restreindre à un seul, et ainsi de construire le sujet en fonction des opportunités.

La découverte du dispositif CIFRE (Conventions Industrielles de Formation par la Recherche) – et de sa déclinaison CIFRE-Défense – est un premier pas.

Créé par le Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation (MESRI) en 1981 et mis en œuvre par l'Association nationale de la Recherche et de la Technologie (ANRT)⁶⁷, ce dispositif « subventionne toute entreprise de droit français qui embauche un doctorant pour le placer au cœur d'une collaboration de recherche avec un

⁶⁶ Master ès Sciences en Sciences criminelles, École des Sciences criminelles, Université de Lausanne obtenu en février 2017.

⁶⁷ Créée en 1953, l'ANRT rassemble les acteurs publics et privés de la recherche et du développement en France. Elle a pour objectif d'aider à améliorer l'efficacité du système français de recherche et d'innovation, notamment en facilitant les relations entre les secteurs publics et privés.

laboratoire public⁶⁸ ». Il a pour objectifs de favoriser le développement de la recherche partenariale publique-privée et d’immerger les doctorant(e)s dans des conditions tant de recherche en laboratoire que d’emploi en entreprise⁶⁹. Ce dispositif repose sur l’association de quatre acteurs – l’entreprise, le laboratoire de recherche académique, la doctorante et l’ANRT – et fonctionne comme suit : la doctorante, diplômée d’un niveau Master ou équivalent, est recrutée en CDI ou en CDD⁷⁰ par l’entreprise ; le laboratoire de recherche académique encadre les travaux de la salariée-doctorante inscrite à ce titre dans l’école doctorale de rattachement du laboratoire ; la doctorante consacre 100% de son temps entre le laboratoire et l’entreprise, avec une répartition établie par les parties et indiquée dans le contrat de collaboration de recherche⁷¹ ; enfin, l’ANRT contracte avec l’entreprise une Convention industrielle de Formation par la Recherche sur la base de laquelle une subvention annuelle, d’une durée non extensive de trois ans, est versée à l’entreprise⁷² par le MESRI (*via* l’intermédiaire de l’ANRT). La demande de CIFRE fait l’objet d’une évaluation de la valeur scientifique et de l’adéquation du profil de la candidate par un expert scientifique mandaté par l’ANRT.

Sa déclinaison CIFRE-Défense, qui « subventionne les entreprises qui embauchent des doctorants travaillant sur des travaux d’intérêt défense en collaboration directe avec des équipes de recherche extérieures à leur structure⁷³ », créée en 2008, fonctionne de manière similaire, à ceci près que l’entreprise bénéficie d’une subvention du Ministère des Armées *via* l’intermédiaire de l’Agence de l’Innovation de Défense (AID)⁷⁴ et s’applique à des thèses d’intérêt pour le secteur de la défense dont l’évaluation est effectuée par un expert mandaté par l’AID.

Ainsi, bien que la démarche n’apparaisse pas simple, l’existence de ce dispositif rend le projet viable : en élaborant une collaboration avec une entreprise et un laboratoire de recherche académique, un financement est possible. Une première question se précise : quelles entreprises

⁶⁸ Clarisse Angelier et Jeanne Courouble, *Ces créateurs d’entreprises innovantes : quand doctorat se conjugue avec entrepreneuriat*, Paris, Eyrolles-Éd. d’Organisation, 2011, 156 p., p. 6.

⁶⁹ *Le dispositif Cifre*, <https://www.anrt.asso.fr/fr/le-dispositif-cifre-7844>, consulté le 28 mai 2022.

⁷⁰ Articles D. 1242-3 & 6 du code du travail.

⁷¹ Ce document, qui doit être remis signé par les parties au plus tard six mois après le début de la CIFRE, stipule les conditions de déroulement du partenariat, en spécifiant notamment le déroulement du partenariat et la méthodologie de recherche, les lieux d’exercice de la doctorante, les questions de confidentialité et de propriété intellectuelle.

⁷² A cette subvention peut s’ajouter un crédit d’impôt recherche (CIR).

⁷³ *Cifre-Défense*, <https://www.anrt.asso.fr/fr/cifre-defense-24859>, consulté le 28 mai 2022.

⁷⁴ Créée le 1^{er} septembre 2018 par Florence Parly, alors Ministre des Armées, l’AID fédère les initiatives d’innovation du ministère et a pour objectif de « doter le ministère des armées d’un dispositif efficace couvrant l’ensemble des domaines de l’innovation et permettre l’émergence de nouvelles formes d’innovations et d’innovateurs. » (Florence Parly, 2018).

en lien avec le secteur de la défense seraient intéressées par un projet de doctorat proposant d'étudier la criminalité organisée ?

Après plusieurs semaines de recherche et de prises de contact avec de nombreuses entreprises, s'ensuit en juillet 2016 la rencontre avec le président fondateur d'une entreprise naissante proposant des conseils et expertises en géopolitique et relations internationales. Titulaire d'un doctorat de géopolitique et spécialisé dans les trafics d'armes internationaux, conscient des apports de la recherche, il est intéressé par l'idée d'accueillir une doctorante travaillant sur des questions liées au crime organisé au sein de son entreprise. À ce moment naît le projet de thèse à proprement parler, qui se précise autour d'une thématique passionnante et bien connue du président-fondateur : le trafic d'armes à feu. Ce tout premier projet de thèse est intitulé « Modélisation des réseaux de circulation d'armes alimentant le terrorisme à travers le monde ». Malheureusement, en juillet 2017, à quelques mois de la date limite de dépôt des candidatures annuelles en CIFRE-Défense auprès de l'AID, après une année de construction du projet, le président-fondateur de cette entreprise décide finalement de ne pas y donner suite, certainement pour des raisons financières.

La recherche d'une entreprise d'accueil reprend à son point initial. La rencontre en septembre 2017 de Didier Bazalgette, alors Responsable Innovation en sciences humaines de la société et intelligence artificielle à l'AID, puis de Christophe Fagot, Responsable R&D chez Intactile Design, redonne un premier souffle à cette ambition de thèse. En effet, Intactile est une agence de design et création d'interfaces numériques concevant notamment un logiciel pour le secteur de la défense, financé en partie sur fonds propres et en partie par l'AID par le biais de projets de recherche. Intéressés par le profil pluridisciplinaire et l'approche de la criminalité organisée par le renseignement criminel, C. Fagot et D. Bazalgette insistent cependant sur le fait que cet intérêt concerne l'analyse de réseaux et non la thématique du trafic d'armes à feu. Ainsi se construit avec Christophe Fagot une ébauche, et progressivement un projet de thèse alors intitulé « l'exploitation de données réelles issues d'un micro-cas à l'aide d'un nouveau moyen d'intégration, d'analyse et de visualisation – le logiciel *Casym*⁷⁵ – peut-elle aboutir à une modélisation de réseaux criminels et appuyer le travail de l'analyste dans la détection de points faibles et la proposition de solutions proactives ? ».

Deux étapes majeures sont encore à accomplir avant de pouvoir soumettre le projet de thèse CIFRE-Défense à l'AID : en premier lieu, rechercher un laboratoire de recherche

⁷⁵ Développé par l'agence Intactile design, le logiciel *Casym* est aujourd'hui commercialisé sous le nom *Stemic*.

académique pour assurer l'encadrement scientifique de la doctorante. En second lieu, obtenir un accès au matériau de travail essentiel à la modélisation systémique de réseaux criminels : des données (opérationnelles, tactiques ou stratégiques) portant sur la criminalité organisée.

En janvier 2018, Sébastien-Yves Laurent, Directeur du Master en Sécurité globale et Analyse Trilingue (SGAT) et Professeur à l'Université de Bordeaux accepte d'encadrer cette thèse de doctorat, sous réserve que le projet se concrétise par un accès à des données consistantes. Travaillant depuis près de vingt ans sur les enjeux de renseignement et possédant une très bonne connaissance des environnements administratifs sensibles, le Professeur Laurent apporte une aide déterminante pour l'accès aux données non seulement en transmettant mais aussi en appuyant le projet de thèse, devenu plus solide et construit et s'étant doté d'une composante de sciences humaines et sociales, auprès de l'Institut de Recherche criminelle de la Gendarmerie nationale (IRCGN).

Le dossier est ensuite envoyé au Service central du Renseignement criminel de la Gendarmerie nationale (SCRC), dans les activités duquel il s'inscrit davantage. En avril 2018, le Lieutenant-Colonel Jérôme Barlatier, Chef de la Division du Renseignement (DR) du SCRC manifeste son intérêt pour le projet et aide à le fortifier en proposant une vision métier et une expertise d'analyse. Lorsque la version finale du projet de thèse est atteinte, en juin 2018, le Lieutenant-Colonel Barlatier porte la demande d'accès à des données de gendarmerie auprès de sa hiérarchie.

En octobre 2018, la Gendarmerie nationale donne son accord verbal pour l'accès aux éléments de procédure d'une enquête à déterminer dans les premiers mois des travaux, sous réserve que le projet de thèse se concrétise effectivement avec les acteurs mentionnés. Ainsi, après plus de deux ans d'élaboration, le projet inclut tous les acteurs nécessaires à sa mise en place, et, grâce à leurs apports respectifs, est suffisamment robuste et élaboré pour être soumis à l'ANRT sous l'intitulé « Appréhension des réseaux criminels par les connaissances sociologiques et expertises des enquêteurs spécialistes : assistance dans l'analyse et les enquêtes sur les réseaux criminels par la modélisation systémique des approches criminelles et sociologiques⁷⁶ ». La date limite annuelle de dépôt des candidatures en thèse CIFRE-Défense étant fixée en septembre, le projet n'a pu être soumis à l'AID à temps, raison pour laquelle il

⁷⁶ Comme précisé en introduction du présent manuscrit, le titre sera amené à évoluer au cours des trois années de recherche.

est finalement soumis à l'ANRT dans le cadre du dispositif CIFRE classique – ce qui n'en modifie en rien la consistance – en octobre 2018.

En novembre 2018, l'inscription en 1^{ère} année de doctorat à l'École doctorale Montaigne Humanités de l'Université de Bordeaux Montaigne est validée. En février 2019, le dossier de candidature en thèse CIFRE est accepté par l'ANRT. Le début officiel des travaux est fixé par accord entre les parties au 15 juin 2019.

I.2 Les acteurs du projet et la construction du sujet

Ce projet tripartite implique ainsi les acteurs suivants : l'Institut de Recherches Montesquieu (IRM – Université de Bordeaux) en tant que laboratoire de recherche rattaché à l'École doctorale Montaigne-Humanités (EDMH – Université Bordeaux Montaigne) ; l'entreprise Intactile Design, qui emploie la doctorante ; et le Service central du Renseignement criminel de la Gendarmerie nationale (SCRC), qui fournit le cas d'usage et apporte une expertise métier.

La thèse de doctorat a pour objet de proposer une méthodologie d'analyse et de modélisation des interactions sociales dans un réseau criminel, afin d'en identifier les acteurs les plus stratégiques, *i.e.* les plus essentiels à son fonctionnement. L'objectif est de permettre de détecter des *patterns* dans les données afin d'identifier les individus occupant les positions et fonctions clefs au sein du réseau étudié sur la base des données fournies par la Gendarmerie nationale.

La recherche se veut fondamentalement pluridisciplinaire, et ce pour deux raisons : premièrement, car les enseignements acquis au cours d'un parcours académique diversifié posent la pluridisciplinarité comme l'art de considérer plusieurs champs académiques comme des outils et vecteurs de solutions se nourrissant mutuellement et de trouver la pièce qui permettra de les axer ensemble pour résoudre des problèmes. À titre d'exemple, l'orientation des unités d'enseignement de la Licence de Sciences et Technologie, Mention Physique, Parcours Physique & Interfaces⁷⁷, considère la physique comme plus qu'une discipline en soi : comme un outil s'appliquant à des domaines aussi variés que la biologie, la géologie, la mécanique, le bâtiment, l'étude des fonds marins, la détermination de causes d'incendies et d'explosion... Les physiciens utilisent les mathématiques comme un outil leur permettant de

⁷⁷ Université Pierre et Marie Curie (Paris VI).

modéliser le monde qu'ils observent, les mathématiciens emploient l'informatique pour augmenter leur puissance de calcul, etc.

Secondement, car le sujet s'est construit en étroite collaboration avec les acteurs de la thèse, partageant un intérêt pour la thématique de recherche, bien que tous spécialisés dans des disciplines différentes. Il s'agit ici d'appréhender et décrypter un réseau criminel par des connaissances sociologiques, l'expertise des enquêteurs et analystes, ainsi que la combinaison de méthodes de modélisation issues de l'analyse de réseaux sociaux (ARS), de l'analyse criminelle, de la sémantique, de la statistique, ou encore de l'informatique. Le travail consiste en l'analyse, d'un point de vue sociologique et sociométrique, de la structure et du fonctionnement du réseau criminel issu du cas d'usage à disposition. Cette hybridation des points de vue, des outils et des méthodes confère à cette thèse tripartite un caractère innovant.

a) L'Institut de Recherche Montesquieu (IRM) : le laboratoire de recherche académique et la direction de thèse

Le laboratoire de recherche en science politique « Institut de Recherches Montesquieu » (IRM) de l'Université de Bordeaux comprend un axe de recherche « études de sécurité » qui s'appuie à la fois sur un programme de recherche pluridisciplinaire (Conflictarm⁷⁸) dans les trois disciplines que sont le droit international, l'économétrie et la science politique et sur le master 2 « Sécurité globale et analyste trilingue » (SGAT), qui forme des analystes en sécurité globale (*i.e.* interne et externe).

Ainsi, l'IRM développe une réflexion sur les outils et méthodes scientifiques, basés sur des analyses statistiques et des modèles informatiques pouvant appuyer les méthodologies actuelles déjà connues et employées par les analystes afin de proposer de nouvelles représentations des théories de sciences sociales développées. Dans ce cadre s'est concrétisé le rattachement du présent travail à l'IRM, pour lequel l'intérêt d'une thèse pluridisciplinaire est d'accompagner et favoriser le couplage des méthodes d'analyses qualitatives (telle qu'actuellement employées en sciences politiques et sciences humaines) à des méthodes quantitatives et/ou outillées pour l'aide à la décision.

La thèse est sous la direction du Professeur Sébastien-Yves Laurent, politiste et historien, professeur à l'Université de Bordeaux (UB). Ancien vice-président à la recherche pour

⁷⁸Le projet « Conflictualité armée : déterminants, qualifications et normes, choix stratégiques et impacts économiques » (Conflictarm) fait suite à un appel à « projets interdisciplinaires SHS » du département Sciences Humaines et Sociales de l'Université de Bordeaux. Porté par les trois disciplines des sciences sociales que sont l'économie, la science politique et le droit, l'intérêt scientifique en est de croiser les différentes approches afin de s'approcher au plus près de la complexité de la problématique étudiée.

les sciences humaines et sociales (2018-2022), il dirige actuellement le laboratoire Institut de recherches Montesquieu (IRM). Au sein de la Faculté de droit et de science politique, il dirige également le master 2 de science politique « Sécurité globale et analyste trilingue » (SGAT) sur lequel s'appuie l'axe de recherche « études de sécurité » du laboratoire. Sébastien-Yves Laurent est diplômé de Sciences-Po Paris et de la Sorbonne, agrégé, docteur et HDR (Sciences-Po Paris). Ses principaux intérêts de recherche, aussi bien en recherche fondamentale ou appliquée, sont : l'épistémologie appliquée, les enjeux de cyber et technologiques dans le système international, les approches pluridisciplinaires de la conflictualité armée et les méthodes analytiques de renseignement, d'anticipation et de prospective. Il dirige une dizaine de thèses de doctorat sur les enjeux de sécurité internationale et de technologie dont plusieurs thèses CIFRE en lien avec l'industrie.

La direction du Professeur Laurent apporte de nombreux éléments conceptuels et une solide base bibliographique, tant par ses connaissances en sciences humaines et sociales que par son expertise sur les questions de renseignement et sa volonté de valorisation de la pluridisciplinarité.

b) Intactile Design, l'entreprise d'accueil

Intactile Design est une agence de design numérique concevant et développant des logiciels pour les milieux professionnels tels que la santé, le contrôle aérien, ou encore la défense. Dans le cadre de ce dernier domaine, Intactile travaille depuis plus de quinze ans sur un environnement d'aide à la réflexion basé sur des modèles de connaissances : l'outil *Stemic*⁷⁹.

Leurs travaux sur les environnements d'aide à la réflexion concernent deux profils d'utilisateurs. Les premiers sont ceux dont l'objectif est de travailler sur la mise à disposition de modèles de connaissances, les seconds sont ceux qui peuvent bénéficier de ces modèles dans leur travail, tels que des enquêteurs ou analystes. Le sujet de thèse intéresse Intactile au premier chef car il a pour objectif de mener des travaux de recherche permettant de fournir des modèles de connaissances travaillés et mis au point par des personnes formées aux métiers de l'analyse criminelle et utilisables dans différents types d'enquêtes, à court et moyen-terme.

Dans le cadre de l'entreprise, la thèse est supervisée par Christophe Fagot, titulaire d'un doctorat en informatique de l'Université des Sciences de Montpellier, avec une spécialisation

⁷⁹ Anciennement *Casym*.

en intelligence artificielle sémantique. Actuellement directeur des opérations de l'agence Intactile Design, il travaille depuis près de vingt ans sur des technologies et des environnements logiciels d'aide à la décision et à la réflexion. Ces travaux, appliqués à des domaines aussi variés que la médecine, le droit, la fiscalité, et la stratégie militaire, aboutissent aujourd'hui à la création et la diffusion de logiciels en rupture avec l'usage actuel de l'intelligence artificielle statistique. En prenant les problématiques par les besoins des usagers et non par les capacités des technologies bien souvent mises au point hors sol, le travail interdisciplinaire unissant des designers et des informaticiens permet d'aller dans le sens de technologies au service de l'humain et non l'inverse. Ainsi, la supervision de C. Fagot représente un apport considérable en termes tant de connaissances que de méthodes et raisonnement sur les aspects informatiques et sémantiques de la modélisation de réseaux et d'analyse de graphes.

c) Le Service central du Renseignement criminel de la Gendarmerie nationale (SCRC) : l'apport des données et de l'expertise métier

Forte de 100.000 militaires, la Gendarmerie nationale assure les missions de sécurité publique, de police judiciaire et de maintien de l'ordre sur 95% du territoire national au profit de 50% de la population. Cette force de sécurité intérieure repose sur une structure pyramidale fortement territorialisée, assurant la complémentarité et la subsidiarité de ses unités généralistes et spécialisées.

Au sein du pôle judiciaire de la Gendarmerie nationale (PJGN), le fonctionnement du Service central de Renseignement criminel (SCRC) repose sur les principes d'une police guidée par le renseignement (*Intelligence-Led Policing - ILP*). Le SCRC regroupe ainsi un ensemble d'expertises concernant le renseignement criminel stratégique sous forme d'aide à la compréhension des phénomènes criminels, d'aide à l'enquête ou d'aide à la décision, mais aussi d'appui opérationnel à l'enquête (analyse criminelle opérationnelle, sciences du comportement, atteintes aux personnes, armement explosif, identification des véhicules, fraudes à l'identité, délinquance économique et financière...).

Afin d'anticiper les évolutions de la criminalité et d'envisager des modes opératoires innovants pour les forces de l'ordre, le SCRC développe également une politique volontariste de recherche et de développement. Cette politique est mise en œuvre au moyen de ses ressources propres en chercheurs, ainsi que d'une politique de partenariat avec des laboratoires. Actuellement, ses principaux axes d'effort sont le développement de systèmes de collecte et

d'analyse de l'information de masse, la mise à profit des évolutions de l'intelligence artificielle et l'exploration des potentialités des objets connectés (*IoT - Internet of Things*).

Cette thèse et les partenariats qu'elle initie intéressent le SCRC dans le cadre de cette politique de recherche, tant au niveau des domaines d'expertise du renseignement que de l'appui à l'enquête. Ses aspects pluridisciplinaires et le développement conjoint d'une méthodologie de gestion et d'analyse de données en lien avec les réseaux criminels s'intègrent bien dans les missions de recherche du SCRC.

Dans le cadre du SCRC, la thèse est supervisée par le Lieutenant-Colonel Jérôme Barlatier, Docteur en criminologie de l'université de Lausanne et chef de la Division du Renseignement (DR) du SCRC. Fort de 22 ans d'expérience professionnelle orientée sur l'exercice de la police judiciaire au sein de la Gendarmerie nationale, il a alterné des fonctions de commandement opérationnel (unité d'observation et de surveillance, division d'investigation en charge de la lutte contre les vols à main armée, chef de circonscription de sécurité publique) et des fonctions d'environnement (en charge du renseignement criminel en administration centrale). Il s'attache à promouvoir une analyse stratégique et opérationnelle tout à la fois ancrée dans la donnée de terrain et ouverte sur le monde de la recherche. Outre l'accès aux données qui constituent le cas d'usage de la recherche, le Ltd-Col. Jérôme Barlatier apporte une solide expérience métier et une forte expertise en analyse criminelle, permettant à cette thèse de s'inscrire concrètement dans la thématique du renseignement criminel.

I.3 Définition des termes du sujet et des théories mises en application

Le titre final de cette thèse entend refléter son caractère pluridisciplinaire et mettre en avant les différents intérêts des acteurs impliqués : *Modélisation hybride de réseaux dans un « champ criminel » : contribution des sciences sociales et d'outils logiciels au renseignement criminel*.

Les paragraphes suivants proposent une définition des termes du sujet dans le cadre de la recherche. Quatre aspects sont à distinguer : en premier lieu le renseignement et l'analyse criminelle ; en deuxième lieu, la notion de criminalité organisée ; en troisième lieu, les notions de champ et de structures sociales, issues des théories du sociologue Pierre Bourdieu ; et enfin la modélisation de ces structures à l'aide d'outils théoriques et logiciels.

Le renseignement est « un processus élaboré de cueillette, de vérification, de classification, d'analyse et de diffusions particulières utiles à l'analyse de sécurité »⁸⁰. Selon Maurice Cusson, le renseignement peut être utilisé d'au moins huit manières. Deux d'entre elles illustrent les objectifs de la recherche : la compréhension d'un fonctionnement ainsi que la connaissance de profil(s) de criminalité organisée. L'appréhension d'un fonctionnement conduit à la recherche de contre-mesures préventives : où et quand, contre qui et par quels moyens agir ? Les informations sur un profil de criminalité, les « points chauds », les principales organisations criminelles opérant sur le territoire et leurs champs d'action sont autant d'atouts pour la planification stratégique telle que l'équilibre entre prévention et répression ou le choix de manœuvres d'opérations ou plans d'actions spéciaux⁸¹.

Pour être menées à bien, ces deux démarches font notamment appel à l'analyse criminelle. Il en existe de nombreuses définitions, nous retiendrons ici celle adoptée par douze pays européens membres d'Interpol en 1992 et sur laquelle d'autres pays se sont depuis alignés : « L'analyse criminelle consiste en la recherche et la mise en évidence méthodiques de relations d'une part entre des données de criminalité elles-mêmes et d'autre part en des données de criminalité et d'autres données significatives, à des fins de pratiques judiciaires et policières »⁸². Le Groupe de Travail sur l'Analyse criminelle (GTAC) d'Interpol explicite que cette démarche emploie des techniques normalisées pour élaborer et vérifier des hypothèses, reconstituer le déroulement des faits, reconnaître et détecter des infractions commises par les mêmes auteurs ou groupes d'auteurs ainsi que pour comprendre le fonctionnement de groupes criminels organisés et étudier les caractéristiques et l'ampleur des activités criminelles⁸³.

Dans ce cadre, le renseignement criminel est défini comme suit par le Lieutenant-Colonel Jérôme Barlatier, Chef de la division renseignement du SCRC : « Le renseignement criminel consiste en l'acquisition d'un savoir en vue de l'action dans un but de prévention et de répression de la criminalité à partir de sources ouvertes ou fermées de toutes natures en recourant à une méthode et à des outils spécifiques »⁸⁴. Il s'agit de « l'ensemble des méthodes et techniques qui orientent le recueil et le traitement des données pertinentes, l'analyse des informations et la production de renseignements qui influencent l'action de sécurité et l'enquête

⁸⁰ Maurice Cusson, Frédéric Lemieux et Benoît Dupont, *Traité de sécurité intérieure*, op. cit., p. 47.

⁸¹ *Ibid.*

⁸² Interpol (Groupe de travail sur l'analyse criminelle), *Guide sur l'analyse criminelle*, Lyon, Interpol, 1997, 66 p., p. 10.

⁸³ *Ibid.*

⁸⁴ Jérôme Barlatier, cours « Présentation du renseignement criminel », stage d'analyse en renseignement criminel, SCRC, 2018.

judiciaire »⁸⁵. Enfin, le modèle de police guidée par le renseignement (*ILP*⁸⁶) est défini par Ratcliffe comme « un modèle organisationnel et une philosophie de management dans lequel l'analyse des données et le renseignement criminel jouent un rôle central dans la prise de décision relative à la réduction de la criminalité et des désordres, et la prévention de la délinquance grâce à la fois à une stratégie d'organisation et à une doctrine judiciaire pragmatique ciblant les délinquants sérieux et récidivistes »⁸⁷.

Si le renseignement criminel s'intéresse à toutes les formes de criminalité, le présent manuscrit se limite au phénomène particulier, bien que multiforme, de la criminalité organisée, véritable fléau social. Ses effets sont ressentis dans le monde entier, à différentes échelles, dans la mesure où la criminalité organisée érode la gouvernance des États, infiltre les processus politiques, alimente la violence et la corruption, exploite les personnes vulnérables et sapent les marchés licites⁸⁸.

Il est cependant intéressant de noter que la « criminalité organisée », ou « criminalité en bande organisée » n'est pas juridiquement définie en France. Cependant, l'article 132-71 du Code pénal explicite le concept de « bande organisée » ainsi : « constitue une bande organisée au sens de la loi tout groupement formé ou toute entente établie en vue de la préparation, caractérisée par un ou plusieurs faits matériels, d'une ou de plusieurs infractions ». La notion même de bande organisée est ainsi dénuée de critères précis tels que le nombre d'individus impliqués ou la nature de la (ou des) infraction(s) commise(s)⁸⁹. Il s'agit ici d'une « circonstance aggravante »⁹⁰, tandis que le Code pénal prévoit comme infraction indépendante « l'association de malfaiteurs » (article 450-1), à la définition quasi identique. Ainsi, malgré l'existence d'une convention internationale visant à définir la criminalité organisée⁹¹, les États maintiennent leurs propres définitions et concepts, limitant ainsi le consensus global⁹².

⁸⁵ Olivier Ribaux et Quentin Rossy, cours « Analyse criminelle opérationnelle », ESC – Université de Lausanne, 2014.

⁸⁶ *Intelligence-Led Policing*.

⁸⁷ Jerry Ratcliffe, *Intelligence-led policing*, London ; New York, Routledge, Taylor & Francis Group, 2016, 222 p., traduction par Clément De Maillard dans Clément De Maillard, « La France et le renseignement criminel : entre volonté et réalité, une ambition à écrire », Sécurité et stratégie, vol. 17 no 2, 2014, p. 49-59, p. 50.

⁸⁸ Office des Nations Unies contre les Drogues et le Crime (ONUDC), *Référentiel stratégique de lutte contre la criminalité organisée pour l'élaboration de stratégies à fort impact*, Vienne, Nations Unies, 2021.

⁸⁹ Salomé Lannier, « Blanchiment de la traite des êtres humains : réflexions générales sur la complémentarité des deux approches », *Cahiers de la sécurité et de la justice*, n° 48-49, 2020, pp. 271.

⁹⁰ Une circonstance aggravante est une situation particulière dans laquelle une infraction est commise et qui confère à celle-ci un caractère de gravité accru entraînant de ce fait l'aggravation de la peine encourue.

⁹¹ Convention des Nations unies contre la criminalité transnationale organisée, 2000.

⁹² Valsamis Mitsilegas, « From National to Global, from Empirical to Legal: The Ambivalent Concept of Transnational Organized Crime », dans Margaret E. Beare (dir.), *Critical Reflections on Transnational Organized Crime, Money Laundering, and Corruption*, Toronto, University of Toronto Press, 2003, pp. 55-87.

Ainsi, nous analyserons le réseau criminel constitué par le cas d'usage au prisme de la notion de « criminalité organisée » du point de vue de la criminologie plutôt que de celui de la jurisprudence, et emprunterons la définition de Carlo Morselli, Mathilde Turcotte et Guillaume Louis : « le crime organisé offre des services publics prohibés dans le but de réaliser des profits privés »⁹³. Cette définition s'accorde avec les interprétations d'un certain nombre de criminologues, tels que Thorsten Sellin⁹⁴ ou George Vold, Thomas Bernard et Jeffrey Snipes⁹⁵, pour qui la notion inclut aussi bien l'offre de biens et services illicites que l'offre illicite de biens et services licites, tels que les systèmes bancaires alternatifs. Selon cette interprétation, les trois auteurs francophones posent pour axiome que « le crime organisé apparaît dans les interstices de la vie économique, politique et sociale soit du fait de la déficience d'un État dans l'approvisionnement de certains services légitimes, soit à la suite d'une interdiction par les institutions de services pour lesquels il existe une forte demande »⁹⁶. Deux thèses majoritaires – et souvent complémentaires – ont émergé de ce principe. D'une part, la thèse des « interstices criminogènes » fait référence aux vides laissés par les institutions dans l'offre de services et dans la rédaction des régulations, créant ainsi des opportunités pour les groupes criminels⁹⁷. D'autre part, la thèse du « marché criminel » considère les groupes criminels comme des entreprises criminelles et étudie leur place dans les secteurs prohibés. Ces deux points de vue ne sont pas antinomiques, notamment dans la mesure où un même réseau criminel peut proposer de manière illicite des services légaux et illégaux. Comme nous le verrons par la suite⁹⁸, l'organisation étudiée entre dans ce cas de figure. Ses services couvraient tant le transfert d'argent de clients particuliers à l'étranger *via* un système financier parallèle (dit de « l'*hawala* ») que le blanchiment massif d'argent pour des clients tels que des réseaux de trafic de stupéfiants. Il est dès lors important de considérer les deux points de vue pendant l'analyse de la structure sociale.

Une fois les termes du sujet délimités, il convient de s'attarder sur les méthodes mises en œuvre dans le cadre de cette recherche.

⁹³ Carlo Morselli, Mathilde Turcotte et Guillaume Louis, « Le crime organisé et les contre-mesures », dans *Traité de sécurité intérieure*, Lausanne, Presses polytechniques et universitaires romandes, 2007, pp. 185-199., p. 185.

⁹⁴ Thorsten Sellin, « Organized Crime: A Business Enterprise », *The Annals of the American Academy of Political and Social Science*, vol. 347, , 1963, pp. 12-19.

⁹⁵ George B. Vold, Thomas J. Bernard et Jeffrey B. Snipes, *Theoretical criminology*, New York, Oxford University Press, 2002, 433 p.

⁹⁶ Carlo Morselli, Mathilde Turcotte et Guillaume Louis, « Le crime organisé et les contre-mesures », *op. cit.*, p. 186.

⁹⁷ Daniel Béland, « Insecurity, Citizenship, and Globalization: The Multiple Faces of State Protection », *Sociological Theory*, vol. 23, n° 1, 2005, pp. 25-41.

⁹⁸ Cf. chapitre III « Le cas d'usage et les données d'étude ».

D'une part, les sciences humaines et sociales, de par leurs connaissances empiriques et leurs méthodes authentiques d'analyse, sont indispensables pour apporter un caractère qualitatif à l'analyse⁹⁹. Les théories du sociologue Pierre Bourdieu sont pertinentes dans ce contexte de recherche pour travailler sur la structure sociale du réseau criminel que constitue le cas d'usage de cette thèse. Nous nous intéresserons en particulier aux concepts d'espaces et de champs sociaux ; l'espace social étant caractérisé comme « un système de positions sociales qui se définissent les unes par rapport aux autres »¹⁰⁰. La distinction de ces positions sociales et leur agencement relatif constituent la structure sociale de cet espace. Outre les positions sociales, Bourdieu montre que la différenciation des activités à l'intérieur d'un espace conduit à la constitution de sous-espaces sociaux appelés « champs sociaux »¹⁰¹ et dotés de règles sociales spécifiques, qu'elles soient explicites ou implicites.

La vision du monde social de Bourdieu, dont s'inspire cette thèse, est adaptée au domaine d'application que constituent les réseaux sociaux criminalisés. Notamment, la notion de « champ criminel », discutée dans le chapitre VIII, s'inscrit dans cette théorie : l'hypothèse est ici qu'il existe un « espace social criminel » constitué par le système des positions sociales des individus impliqués dans des activités criminelles. Dans cet espace social « criminel » se distinguent différents « champs criminels » régis par des règles sociales spécifiques au type de structure et au domaine d'activité criminelle. Ce n'est donc pas une étude des théories bourdieusiennes que nous réalisons ici, mais plutôt une transposition de leurs concepts à notre cas d'usage.

D'autre part, l'analyse est couplée avec une approche par la théorie des graphes, qui utilise des outils quantitatifs distincts de ceux usuellement appliqués en sociologie bourdieusienne. Les deux approches, qualitatives et quantitatives, sont complémentaires afin d'obtenir la vision la plus exhaustive des réseaux criminels étudiés. En effet, afin de vérifier notre hypothèse, nous appliquerons des méthodes et mesures issues de l'analyse de réseaux sociaux (ARS) – discipline qui rassemble l'ensemble des outils théoriques et méthodologiques permettant l'analyse de données relationnelles et la détection des régularités qu'elles présentent¹⁰². L'ARS s'intéresse aux relations entre entités sociales (individus, groupes d'individus, entreprises, organisations, États, etc.) et permet d'exploiter les informations selon

⁹⁹ Sébastien-Yves Laurent, « L'analyse, la ressource cachée du renseignement », *Revue des deux mondes*, 2016, pp. 60-66.

¹⁰⁰ Alain Accardo, *Introduction à une sociologie critique : lire Pierre Bourdieu*, Marseille, Agone, 2006, 382 p., p.56.

¹⁰¹ Pierre Bourdieu, « Espace social et genèse des "classes" », *op. cit.*

¹⁰² Rémi Boivin et Carlo Morselli, *Les réseaux criminels*, Montréal, Les Presses de l'Université de Montréal, 2016, 319 p.

trois approches : les entités, les relations, et les réseaux comme un ensemble d'entités et de relations. Notre démarche s'effectue notamment au travers de la théorie des graphes, *i.e.* la discipline mathématique et informatique qui se consacre à leur étude. Un graphe est défini comme une représentation de relations entre des entités¹⁰³. Il est composé de deux ensembles d'objets : les nœuds, qui représentent les entités considérées, et les liens, qui illustrent les relations existantes entre ces dernières. Les algorithmes élaborés en théorie des graphes ont ainsi de nombreuses applications en ARS, et seront ici employés pour la modélisation du réseau criminel étudié.

Pour mener à bien l'analyse de la structure du réseau social de l'organisation criminelle étudiée, ce travail s'appuie sur des outils informatiques : des logiciels tels que *Stemic* ou *RStudio* et la rédaction d'algorithmes spécifiques, selon les besoins. Ces outils sont employés pour la visualisation, la représentation, la modélisation et l'analyse de l'ensemble des données à disposition. Le terme « logiciel » est ici employé au sens large, c'est-à-dire comme « un ensemble des programmes, procédés et règles relatifs au fonctionnement d'un ensemble de gestion et traitement de données »¹⁰⁴. L'utilisation de ces outils permet de réaliser des calculs et traitements sur des données en grande quantité. L'objectif est de faciliter la détection et la création de renseignements exploitables du point de vue du renseignement criminel sur la base des données : il s'agit d'assister – et non de remplacer – le travail de l'analyste dont le rôle est, et reste, essentiel dans la prévention et la répression des infractions et des menaces à l'ordre et la sécurité publique¹⁰⁵.

¹⁰³ Stanley Wasserman et Katherine Faust, *Social Network Analysis Methods and Applications*, *op. cit.*

¹⁰⁴ *Le grand Larousse illustré*, Logiciel, Paris, Larousse, 2019.

¹⁰⁵ Jérôme Barlatier, *Management de l'enquête et ingénierie judiciaire, recherche relative à l'évaluation des processus d'investigation criminelle*, Thèse de doctorat, École des Sciences Criminelles, Université de Lausanne, 2017, 475 p.

I.4 Le cycle du renseignement : de la donnée brute au renseignement

« En amont de la formation d'un renseignement préexiste forcément une réalité politique, sociale, économique ou scientifique qui, avant de devenir une information, peut encore subir différentes transformations. »¹⁰⁶

Sébastien-Yves Laurent.

Le modèle d'*ILP*, né d'une volonté des forces de sécurité de prévenir le crime plutôt que d'y réagir après les faits, peut être synthétisé par le modèle des « 4P » de Ratcliffe : la prévention nécessite de mettre en place des actions proactives, rendues possibles par une forme de prédictibilité mise en évidence par la détection de *patterns*¹⁰⁷, comme illustré sur la figure 2.

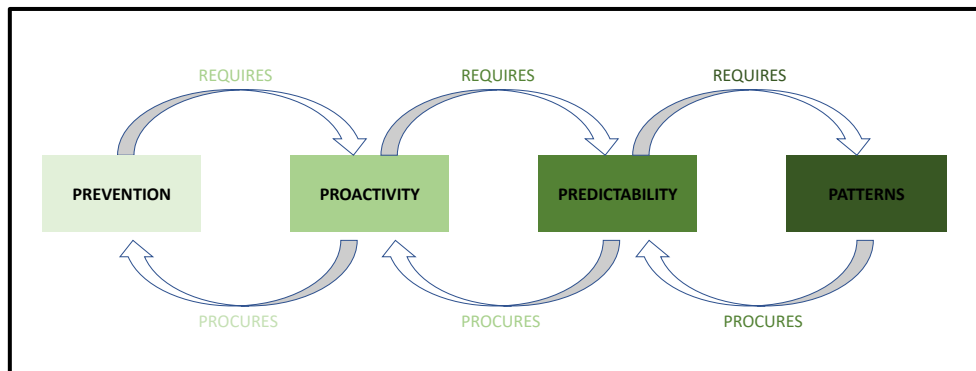


Figure 2 Modèle des « 4P » de Ratcliffe¹⁰⁸

La détection de *patterns* – qu'il s'agisse de *patterns* dans les données (répétitions observées dans les données) ou de *patterns* d'activités (régularités existantes dans la criminalité) – s'inscrit dans le cycle du renseignement. Puissant outil méthodologique régissant le processus de conversion de la donnée en renseignement, l'une des plus synthétiques conceptualisations formelles en est le modèle du *DIKI* (*Data-Information-Knowledge-Intelligence*) de Ratcliffe : la donnée est collectée, c'est-à-dire à disposition de l'analyste qui doit alors la reconnaître comme pertinente et la traiter en la standardisant, l'ordonnant et l'organisant, pour en obtenir une information (figure 3). Cette information est ensuite analysée et interprétée dans son contexte pour en extraire une connaissance. Enfin, lorsque cette

¹⁰⁶ Sébastien-Yves Laurent, *Atlas du renseignement. Géopolitique du pouvoir*, Paris, Les Presses de science Po, 2014, 190 p., p.11.

¹⁰⁷ Jerry H. Ratcliffe, « Intelligence-led policing: Anticipating risk and influencing action », dans *Criminal intelligence for the 21st century*, IALEIA, 2011, pp. 206-220.

¹⁰⁸ Inspiré de Jerry Ratcliffe (dir.), *Strategic thinking in criminal intelligence*, Sydney, Federation Press, 2009, 273 p., p. 2, avec l'aimable autorisation de l'auteur.

connaissance est diffusée et qu'elle donne lieu à une prise de décision – que cette dernière consiste à agir ou non¹⁰⁹, elle devient un ou des renseignement(s).

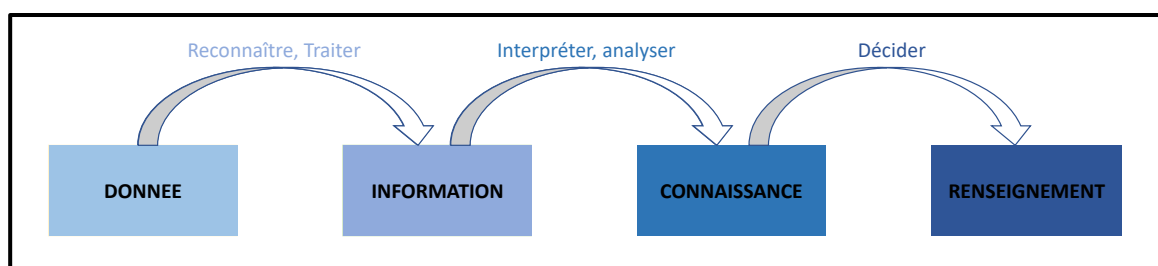


Figure 3 Modèle du *DIKI* de Ratcliffe : de la donnée au renseignement¹¹⁰

Le modèle du *DIKI* schématise cette démarche de manière linéaire afin de représenter la transformation de l'élément « donnée » en une entité « renseignement ». Lorsqu'il s'agit de « produire du renseignement », la démarche est dynamique et itérative : le besoin, les données dans leur ensemble, leur contexte, la situation dans laquelle elles sont évaluées, les outils et méthodes de traitement et d'analyse sont évolutifs dans le temps – et l'espace considéré (figure 4). Ainsi, la collecte de nouveaux éléments, un changement d'environnement ou d'approche amènent à actualiser, enrichir, valider ou réfuter les informations, connaissances et renseignement(s) acquis en premier lieu en procédant à une nouvelle itération du cycle.

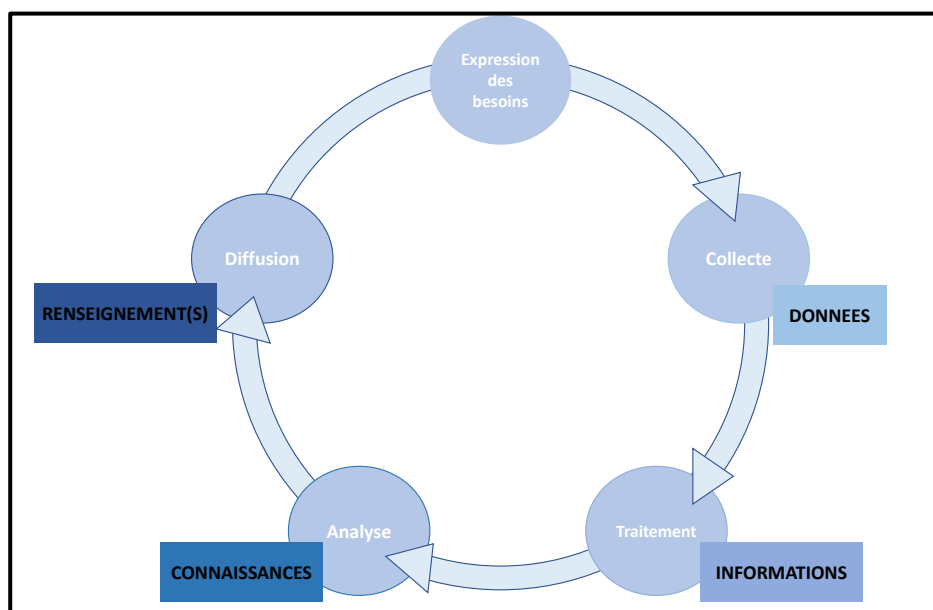


Figure 4 Le cycle du renseignement¹¹¹

L'expression du besoin est le préambule du cycle du renseignement : elle est essentielle pour orienter, coordonner et organiser la recherche et le recueil des données. La compréhension

¹⁰⁹ Sébastien-Yves Laurent, *Atlas du renseignement. Géopolitique du pouvoir*, *op. cit.*

¹¹⁰ Inspiré d'Olivier Ribaux et Quentin Rossy, cours « Analyse criminelle opérationnelle », *op. cit.*, adapté de Jerry Ratcliffe, *Intelligence-Led Policing*, *op. cit.*, p. 72, avec l'aimable autorisation des auteurs.

¹¹¹ Inspiré de Jérôme Barlatier, cours « Présentation du renseignement criminel », *op. cit.*, avec l'aimable autorisation de l'auteur.

du besoin et de ses objectifs et implications (ponctuelle, à moyen ou à long terme et au niveau opérationnel, tactique ou stratégique) est primordiale pour estimer l'étendue du sujet, l'intensité de la menace, l'ampleur des conséquences et le niveau de confidentialité¹¹². Sont ensuite effectuées la recherche et la collecte de données, en sources ouvertes et/ou fermées à l'aide de moyens humains (renseignement d'origine humaine « ROHUM ») ou de moyens technologiques (renseignement d'origine électromagnétique « ROEM », renseignement d'origine image « ROIM »)¹¹³. Cette phase débute par un bilan de l'existant et une identification des données nécessaires et des *intelligence gaps*¹¹⁴ afin d'établir un plan de collecte et un projet d'analyse avant de procéder à l'acquisition, l'enrichissement et la saisie des données. À ce stade, les éléments pertinents ne sont pas encore identifiés : la collecte est volontairement large. L'analyste effectue par la suite un ensemble de traitements sur la donnée afin de la rendre exploitable pour la phase d'analyse : elle doit être nettoyée, classifiée, qualifiée (fiabilité de la source et de l'information elle-même), uniformisée, rendue compréhensible, ordonnée et structurée pour être visualisée¹¹⁵. Enfin, l'analyste s'assure de sa qualité, de sa fiabilité et de sa véracité en mettant en œuvre un processus de vérification afin de minimiser le risque de déformation ou de désinformation¹¹⁶. Une fois traitée, l'information ne doit pas seulement être décrite, mais aussi et surtout être analysée, c'est-à-dire mise en perspective pour refléter les dynamiques temporelles et causales du phénomène étudié¹¹⁷. « Analyser, c'est comprendre et rendre compréhensible l'information par la mise en évidence de liens entre les informations, de *patterns*, de causes et effets¹¹⁸. ».

S'il est recommandé de suivre une méthode bien définie, il est aussi important d'écouter ses idées, et de tester. Il ne s'agit pas, lors de cette phase, d'aboutir à une description parfaite du phénomène, mais bien d'arriver à une compréhension de ce dernier et de ses mécanismes suffisamment détaillée pour permettre de proposer des actions concrètes¹¹⁹. Une analyse de données sur un phénomène doit permettre de refléter et d'expliquer ce dernier. Enfin, la diffusion permet l'exploitation de la connaissance ainsi acquise. Elle peut s'effectuer en interne pour orienter la recherche d'autres renseignements, la prise de décisions et la mise en place

¹¹² Jérôme Barlatier, cours « Expression des besoins », stage d'analyse en renseignement criminel, SCRC, 2018.

¹¹³ Sébastien-Yves Laurent, *Atlas du renseignement. Géopolitique du pouvoir*, op. cit.

¹¹⁴ Lacunes en matière de renseignement.

¹¹⁵ Jérôme Barlatier, cours « Traitement de l'information criminelle », stage d'analyse en renseignement criminel, SCRC, 2018.

¹¹⁶ Sébastien-Yves Laurent, *Atlas du renseignement. Géopolitique du pouvoir*, op. cit.

¹¹⁷ Sébastien-Yves Laurent, « L'analyse, la ressource cachée du renseignement », op. cit.

¹¹⁸ Jérôme Barlatier, cours « Exploration des informations », stage d'analyse en renseignement criminel, SCRC, 2018, p. 5.

¹¹⁹ *Ibid.*

d'actions, ou en externe¹²⁰, afin d'être exploitée par d'autres services ou de faciliter la coopération intra-service ou internationale.

Dans cette thèse, sont réalisées les phases d'expression du besoin (notamment au cours des trois années qui ont séparé l'idée initiale du début des travaux), de traitement, d'analyse, et, à moindre mesure, de diffusion du renseignement. Ce travail de recherche étant effectué sur des données d'enquête de gendarmerie, la phase de collecte a été réalisée en amont de ces travaux par les enquêteurs et les analystes de la gendarmerie ayant travaillé sur l'affaire. Elles ont été récoltées et saisies dans un but d'investigation afin d'avérer des faits et d'identifier des auteurs, ce qui n'est pas l'objet du renseignement criminel. Pour autant, l'enquête judiciaire, de par les données concrètes qu'elle accumule et au travers de la veille opérationnelle, informe sur le fonctionnement des phénomènes criminels qui en font l'objet et nourrit ainsi le renseignement¹²¹ : un phénomène criminel produit des conséquences (humaines, matérielles, financières, écologiques, etc.) et impacte son environnement (de manière directe ou indirecte, à l'échelle individuelle ou sociétale). Cela laisse des traces (matérielles ou immatérielles), qui, une fois détectées, reconnues et récoltées deviennent des données. Ratcliffe explicite cette idée dans sa première loi du renseignement criminel : « *The most reliable indication of future criminal activity is current criminal activity* »¹²² .

Une fois les données analysées au prisme du cycle du renseignement, elles peuvent aboutir à la mise en œuvre d'actions, qui à leur tour impactent ces phénomènes (figure 5). Les renseignements obtenus et l'évaluation de l'effet de ces actions permettent, au fil des itérations du cycle du renseignement avec des données actualisées, d'atteindre une compréhension des mécanismes criminels de plus en plus précise.

¹²⁰ Sébastien-Yves Laurent, *Atlas du renseignement. Géopolitique du pouvoir*, op. cit.

¹²¹ Olivier Ribaux, *Police scientifique*, op. cit.

¹²² Jerry H. Ratcliffe, « Intelligence-led policing: Anticipating risk and influencing action », op. cit., p. 2.

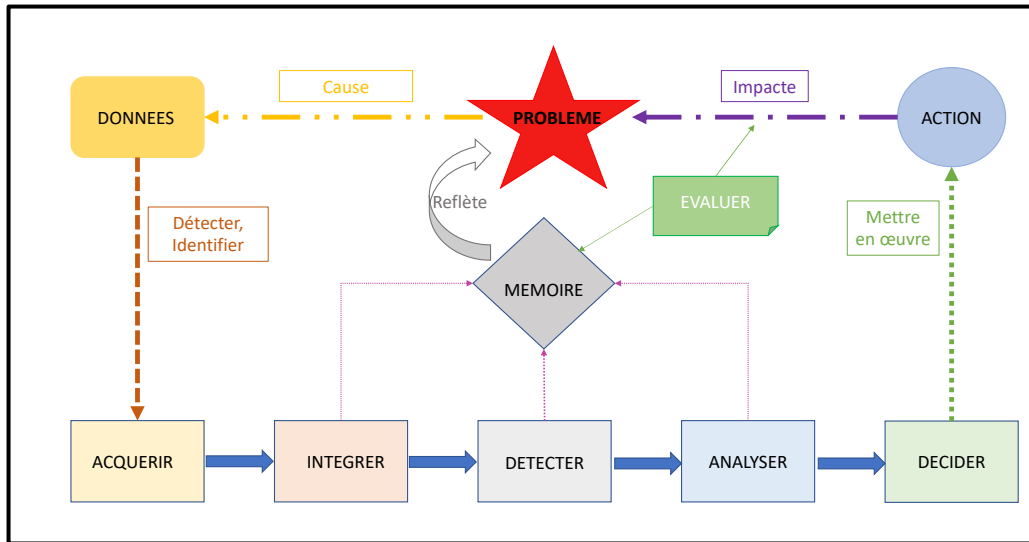


Figure 5 Processus de veille opérationnelle¹²³

« All analysis, even descriptive work of past criminal activity, makes an implicit projection of a possible future criminal environment. »¹²⁴

Jerry H. Ratcliffe.

¹²³ Inspiré d'Olivier Ribaux, *Police Scientifique*, *op. cit.*, p. 296, avec l'aimable autorisation de l'auteur.

¹²⁴ Jerry H. Ratcliffe, « Intelligence-led policing: Anticipating risk and influencing action », *op. cit.*, p. 1.

CHAPITRE II ÉTAT D'AVANCEMENT DES CONNAISSANCES : DES SCIENCES SOCIALES À LA MODÉLISATION

Cette thèse et son sujet, dont l'évolution a été continue tout au long de leur élaboration, se sont construits en étroite collaboration avec les parties prenantes¹²⁵. Le travail de recherche se veut fondamentalement pluridisciplinaire, comme le formalise sa problématique : « Peut-on combiner les outils analytiques issus de multiples disciplines pour développer, sur la base de données telles que collectées et saisies lors des enquêtes de gendarmerie, une méthodologie de représentation, de modélisation et d'analyse de réseaux criminels aboutissant à une détection des fonctions et positions stratégiques dans le fonctionnement de l'organisation ? ». Ce second chapitre est ainsi consacré à l'état de l'art des différentes disciplines, et en particulier des différents concepts méthodologiques, qui seront employés dans les chapitres suivants.

Une première section s'intéresse aux théories issues des sciences sociales abordées. Une deuxième section présente les notions d'analyse des réseaux sociaux (ARS) que nous chercherons à appliquer. La troisième section résume (de manière non exhaustive) l'état des connaissances générales et des méthodes de modélisation en matière d'analyse de réseaux criminels. Enfin, une quatrième section expose les principes de modélisation empruntés aux mathématiques, à l'informatique et aux statistiques lors des phases de modélisation du réseau criminel constitué par le cas d'usage.

II.1 Théories issues des sciences sociales

a) Le concept d'espace social

La littérature en sciences humaines et sociales apporte de nombreux éléments permettant d'étudier les réseaux sociaux. Or, les réseaux criminels sont avant tout des systèmes d'interaction entre individus (appelés « acteurs »), *i.e.* des réseaux sociaux.

Le sociologue Pierre Bourdieu propose une approche en termes d'espace social pour analyser la position – relative – des individus et leurs relations à l'intérieur d'un groupe social. « Un espace social est un système de positions sociales qui se définissent les unes par rapport

¹²⁵ Cf. chapitre I, sections I.1 « Historique de la recherche : du projet au doctorat » et I.2 « Les acteurs du projet et la construction du sujet ».

aux autres »¹²⁶. Les individus sont alors positionnés dans cet espace en fonction des ressources dont ils disposent, les capitaux, et des échanges qui en résultent¹²⁷. Dans sa théorie, Pierre Bourdieu définit les champs comme des sphères de l'espace social, qui, grâce au processus de différenciation progressive du monde social et des activités sociales, se sont autonomisées. Dans un espace social, la différenciation des activités sociales conduit à la constitution de « sous-espaces sociaux » : les champs sociaux.

Ainsi, nous faisons, dans cette thèse, l'hypothèse et l'expérimentation qu'un réseau criminel pratiquant des activités criminelles particulières (telles que le trafic de stupéfiants) et jouissant d'un fonctionnement autonome peut être interprété comme un groupe social – dans lequel évoluent des acteurs – inscrit dans le « champ criminel » spécifique à ses activités (ici, le « champ criminel » du trafic de stupéfiants). Le « champ criminel » s'insère alors dans ce que nous proposons d'appeler « l'espace social criminel », c'est-à-dire le système de positions sociales des individus impliqués dans des activités criminelles (figure 6). En effet, « l'espace social criminel » est régi par des normes et fonctionnements différents du système légal, ce qui l'en distingue. L'un des exemples les plus flagrants en est la notion de justice : il n'existe pas, à l'échelle criminelle, d'appareil judiciaire tel qu'il existe, sous différentes formes, dans le milieu légal.

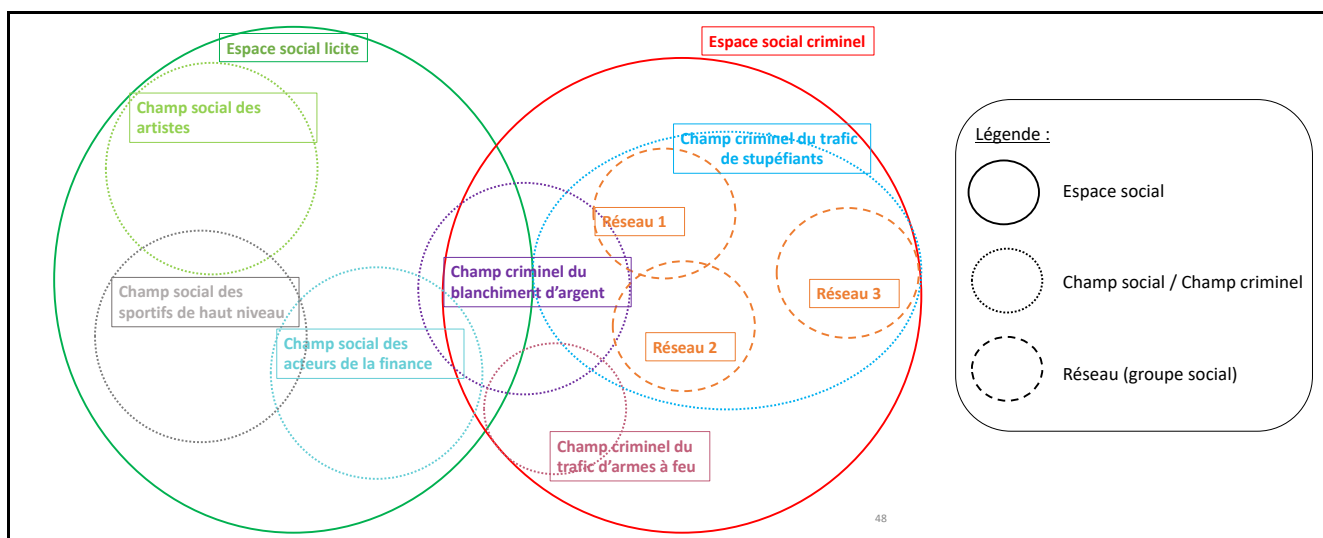


Figure 6 Schéma illustrant l'hypothèse des champs et espaces criminels.

Dans un groupe social, ou à plus grande échelle, dans un champ social (qu'il soit criminel ou non), les acteurs, occupent des positions sociales établies, séparées par des distances sociales. La notion de position sociale se rapporte alors aux positions relatives (et non absolues)

¹²⁶ Alain Accardo, *Introduction à une sociologie critique, op. cit.*, p.56

¹²⁷ Pierre Bourdieu, « Le capital social : Notes provisoires », *Actes de la recherche en sciences sociales*, vol. 31, n° 1, 1980, pp. 2-3.

des acteurs les uns par rapport aux autres dans le groupe social étudié. En effet, si les positions absolues des acteurs, ou fonctions¹²⁸ dans le réseau, doivent être identifiées, seule une compréhension de l'organisation des différentes positions sociales les unes par rapport aux autres permettra de cerner les mécanismes sous-jacents au fonctionnement du réseau dans sa globalité, *i.e.* en tant qu'entité, et ses spécificités, *i.e.* en tant qu'acteur d'un « champ criminel », sous-ensemble singulier de ce que nous avons appelé plus haut « l'espace social criminel ».

La position absolue d'un acteur se rapporte à sa fonction, ou rôle *de facto* dans l'organisation, ce qui s'illustre assez bien dans les systèmes organisationnels bien connus. Par exemple, dans l'armée, l'organigramme du commandement permet de comprendre immédiatement la fonction occupée par chacun. La position relative, en revanche, sera un indicateur du fonctionnement du système social que constitue l'organisation, et des échanges qui y ont lieu. Pour reprendre le même exemple, afin d'anticiper l' (ou les) acteur(s) le(s) plus susceptible(s) d'occuper une position donnée, il est nécessaire d'étudier les échanges et liens sociaux entre les différents acteurs de l'organigramme, et d'en comprendre la dynamique sociale.

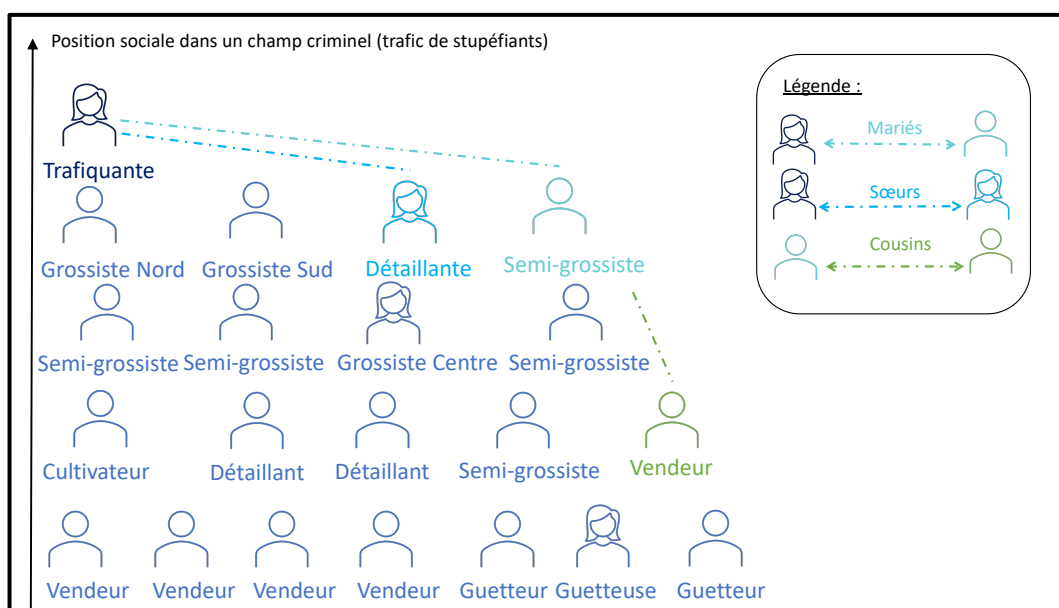


Figure 7 Illustration d'un système de positions absolues (les fonctions dans le réseau) et de positions sociales (place des acteurs les uns par rapport aux autres) dans un réseau de trafic de stupéfiants.

La figure 7 représente un réseau imaginé¹²⁹ de trafic de stupéfiants. Les positions sociales des différents acteurs peuvent illustrer leurs niveaux de pouvoir, d'influence,

¹²⁸ Est ici entendu par « fonction », le rôle absolu (livreur, guetteur, comptable...) que l'acteur occupe dans le réseau, à différencier de la position (s'apparentant plus à la notion de « statut », bien que comportant des nuances), qui fait référence au positionnement relatif de l'acteur vis-à-vis des autres agents du champ étudié.

¹²⁹ C'est-à-dire que ce schéma ne représente pas une réalité criminelle, mais est donné à titre d'exemple.

d'autorité, de contacts, *etc.* Dans cet exemple, la hiérarchie sociale, *i.e.* le système des positions sociales, semble dépendre des fonctions des acteurs : la trafiquante a la position sociale la plus élevée, viennent ensuite les grossistes, les semi-grossistes, les détaillants et le cultivateur et enfin les vendeurs et guetteurs. Cependant, certains individus contredisent cette première impression : une détaillante est deux niveaux au-dessus des autres membres occupant sa fonction ; un grossiste et un vendeur sont un niveau au-dessus de leurs homologues. Cela s'explique, dans ce cas, par le fait que ces trois protagonistes entretiennent un lien familial avec la trafiquante qui occupe la plus haute position sociale, qui les fait ainsi bénéficier de son rang privilégié. Ainsi, bien que la détaillante gère en pratique la distribution de plus faibles quantités d'argent et la collecte de montants d'argents plus faibles que les semi-grossistes, sa relation fraternelle lui confère plus d'autorité. D'autres raisons pourraient aussi expliquer ce décalage entre fonction et position : la grossiste du centre peut être très compétente dans cette fonction, mais être moins bien implantée dans le réseau (avoir moins de contacts, d'influence) que ses homologues qui gèrent la distribution dans les régions Nord et Sud.

Ainsi, l'étude des positions et fonctions occupées par les acteurs dans le réseau permet de mieux comprendre les dynamiques sociales à l'œuvre. La compréhension de ces mécanismes est précieuse pour identifier les individus essentiels au fonctionnement du réseau. Les fonctions et positions clefs varient d'un réseau à l'autre, ainsi que d'un champ à l'autre. La figure 8 montre une situation sociale différente dans le « champ criminel » du trafic de stupéfiants.

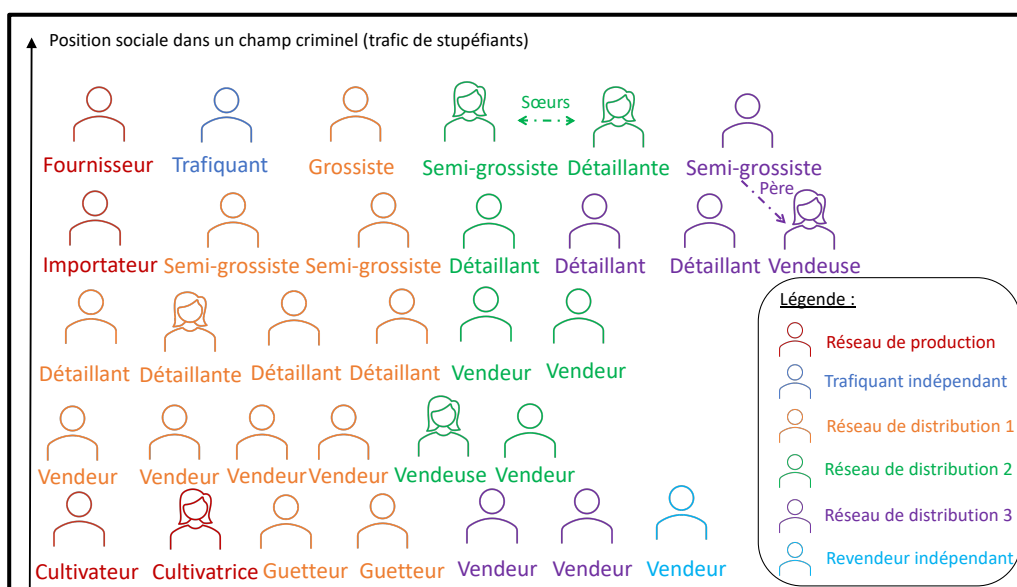


Figure 8 Illustration d'un système de positions absolues (les fonctions dans les réseaux) et de positions sociales (place des acteurs les uns par rapport aux autres) dans une chaîne d'approvisionnement et de distribution de produits stupéfiants.

Cette seconde situation imaginée ne représente pas un unique réseau, mais plusieurs acteurs d'une chaîne de distribution : un réseau produisant et important les produits stupéfiants, un trafiquant indépendant dont les clients sont trois réseaux de distribution (de tailles variables)

et un vendeur indépendant. Dans cet exemple, la hiérarchie sociale dépend moins de la fonction absolue dans la chaîne de distribution que de la place des acteurs dans leurs réseaux respectifs : les vendeurs du réseau de distribution vert ont une position sociale plus élevée que leurs homologues exerçant la même fonction dans le réseau violet – hormis la vendeuse, dont la position est plus haute dans la hiérarchie sociale de par son lien de parenté avec le semi-grossiste, dont la position sociale est la plus élevée.

Ces éléments interrogent sur les attributs qui confèrent ou non une position sociale élevée. Une seconde notion devient alors particulièrement informative : le capital. Les capitaux sont définis par Pierre Bourdieu comme « un travail accumulé qui, lorsqu’obtenu de manière exclusive, par des agents ou des groupes d’agents, leurs permettent d’obtenir une énergie sociale »¹³⁰. En d’autres termes, les capitaux sont des propriétés sociales auxquelles les acteurs de l’espace social accordent de la valeur, et que l’analyse sociologique doit établir dans le champ social qu’elle étudie¹³¹.

Dans un champ social donné, le capital dont dispose chaque acteur régit les échanges selon la loi de l’offre et de la demande. La notion de « travail accumulé » renvoie au fait que l’acteur agit constamment, et donc travaille, non seulement pour l’acquérir, mais aussi pour l’entretenir. Ce capital est décliné en quatre types de ressources : économiques, culturelles, sociales et symboliques¹³² (figure 9).

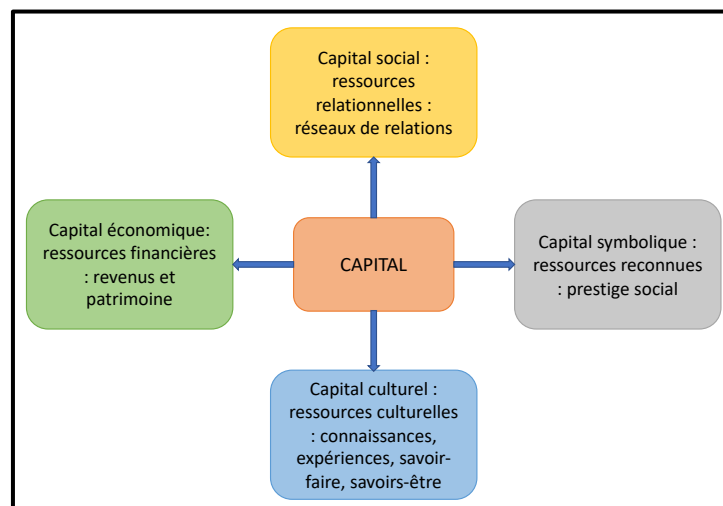


Figure 9 Décliné du capital selon Bourdieu

¹³⁰ Pierre Bourdieu, « The forms of Capitals », dans *Handbook of theory and research for the sociology of education*, Greenwood Press, 1986, pp. 241-258, p. 241.

¹³¹ Gisèle Sapiro, François Denord, Julien Duval, Mathieu Hauchecorne, Johan Heilbron, Franck Poupeau et Hélène Seiler (dir.), *Dictionnaire international Bourdieu*, Paris, CNRS éditions, 2020, 963 p.

¹³² Pierre Bourdieu, « The forms of Capitals », *op. cit.*

Après avoir identifié les fonctions (absolues) des agents, il convient de définir, pour les fonctions-clefs, les capitaux possédés, nécessaires à l'obtention et au maintien de ces rôles par les acteurs qui les occupent. Il devient alors possible de mieux cerner les positions (relatives) des acteurs, et donc leurs interactions (nature, importance stratégique...). Ainsi, le « marché social » et les « flux sociaux » pourraient être modélisés.

Nous nous intéressons en particulier à la compréhension des mécanismes d'échange de ces quatre capitaux dans l'espace social criminel, puis plus spécifiquement au champ criminel d'intérêt, et enfin à l'échelle du groupe social que constitue le réseau criminel étudié. En effet, selon Yanick Charrette¹³³, il n'est pas pertinent d'envisager la notion de capital criminel, dans la mesure où ce dernier peut être modélisé par différentes combinaisons des quatre capitaux bourdieusiens, définis ci-après. Cette idée sera discutée dans le chapitre VIII.

1. Le capital économique

Le capital économique fait référence aux actifs et passifs financiers qu'un agent possède, l'ensemble des biens économiques : revenus, patrimoine et biens matériels. Le fait de disposer de capital économique permet d'acquérir plus facilement d'autres types de capitaux¹³⁴. Pierre Bourdieu l'envisage comme « un pouvoir sur le monde social qui a la particularité de reposer sur la possession de ressources matérielles, extérieures à leur détenteur et d'être, de ce fait, plus facilement transmissibles que les autres formes de capitaux »¹³⁵.

2. Le capital culturel

Le capital culturel¹³⁶ représente les acquis culturels qu'un agent possède, les ressources permettant d'apprécier les biens et les pratiques propres à la culture au sens large. Pierre Bourdieu en résume le concept ainsi : « Le capital culturel recouvre l'ensemble des savoirs, des savoir-faire et des structures de perception et d'appréciation de [ces derniers] »¹³⁷. Le capital culturel existe sous trois formes : à l'état incorporé, *i.e.* sous la forme de dispositions, de savoirs

¹³³ Yanick Charrette est chercheur en criminologie au Centre International de Criminologie Comparée (CICC) de Montréal, professeur à l'École de travail social et de criminologie de l'Université Laval, titulaire d'un bachelors, d'un master et d'un doctorat en criminologie obtenus à l'Université de Montréal, et a effectué un post-doctorat en sociologie à l'Université de Yale.

Entretien réalisé avec Monsieur Y. Charrette par visioconférence le 09/11/2020.

¹³⁴ Anne Jourdain et Sidonie Naulin, *La sociologie de Pierre Bourdieu*, Malakoff, Armand Colin, 2019, 672 p.

¹³⁵ Gisèle Sapiro, François Denord, Julien Duval, Mathieu Hauchecorne, Johan Heilbron, Franck Poupeau et Hélène Seiler (dir.), *Dictionnaire international Bourdieu*, *op. cit.*, p. 109.

¹³⁶ À ne pas confondre avec l'habitus, qui fait référence entre autres à la culture en termes de « valeurs » ou de « morale ».

¹³⁷ Gisèle Sapiro, François Denord, Julien Duval, Mathieu Hauchecorne, Johan Heilbron, Franck Poupeau et Hélène Seiler (dir.), *Dictionnaire international Bourdieu*, *op. cit.*, p. 107.

et savoir-faire, de compétences transmises ; à l'état objectivé, sous la forme de biens culturels ; et à l'état institutionnalisé, sous la forme de titres scolaires¹³⁸.

3. Le capital social

Le capital social est défini comme « l'ensemble des ressources actuelles ou potentielles qui sont liées à la possession d'un réseau durable de relations plus ou moins institutionnalisées d'interconnaissance et d'inter-reconnaissance »¹³⁹. Un agent dispose d'un capital social d'autant plus important que l'étendue de son réseau de relations est grande et que les personnes le constituant sont elles-mêmes fortement dotées en capitaux économiques, culturels et symboliques. Le capital social a un effet multiplicateur vis-à-vis des autres capitaux¹⁴⁰.

4. Le capital symbolique

Le capital symbolique renvoie aux notions de prestige et de reconnaissance sociale. Selon Hegel, la reconnaissance de la particularité et de la singularité de chacun est le lien organique entre une forme de société et sa conceptualisation comme « une combinaison intime de socialisation et d'individualisation »¹⁴¹. Il rapportera par ailleurs dans la suite de ses travaux l'apparition du crime à un défaut de reconnaissance sociale : l'individu qui ne se trouverait pas suffisamment reconnu par ses pairs puiserait dans cette absence son mobile criminel¹⁴².

Pierre Bourdieu élabore initialement le capital symbolique comme une notion renvoyant « aux profits que l'appartenance à une lignée ou à un groupe procure, en particulier le prestige et le renom »¹⁴³. En 1987, il étend ce concept en écrivant que « le capital symbolique n'est autre que le capital économique ou culturel lorsqu'il est connu et reconnu¹⁴⁴ », et ajoute que « [De ce fait,] lorsque [le capital symbolique] est connu selon les catégories de perception qu'il impose, les rapports de force symbolique tendent à reproduire et à renforcer les rapports de force qui constituent la structure de l'espace social ¹⁴⁵ ». Patrice Bonnewitz résume celui-ci comme étant « l'ensemble des rituels (comme l'étiquette ou le protocole) liés à l'honneur et à la

¹³⁸ Pierre Bourdieu, « Les trois états du capital culturel », *Actes de la Recherche en Sciences Sociales*, vol. 30, 1979, pp. 3-6.

¹³⁹ Pierre Bourdieu, « Le capital social : Notes provisoires », *op. cit.*, p. 2.

¹⁴⁰ Anne Jourdain et Sidonie Naulin, *La sociologie de Pierre Bourdieu, op. cit.* ; Gisèle Sapiro, François Denord, Julien Duval, Mathieu Hauchecorne, Johan Heilbron, Franck Poupeau et Hélène Seiler (dir.), *Dictionnaire international Bourdieu, op. cit.*

¹⁴¹ Axel Honneth et Pierre Rusch, *La lutte pour la reconnaissance*, Paris, Gallimard, 2015, 352 p., p. 31.

¹⁴² *Ibid.*

¹⁴³ Gisèle Sapiro, François Denord, Julien Duval, Mathieu Hauchecorne, Johan Heilbron, Franck Poupeau et Hélène Seiler (dir.), *Dictionnaire international Bourdieu, op. cit.*, p. 114.

¹⁴⁴ Pierre Bourdieu, *Choses dites*, Paris, Éditions de Minuit, 1987, 229 p., p. 160.

¹⁴⁵ *Ibid.*

reconnaissance. [...] Il permet de comprendre que les multiples manifestations du code de l'honneur et des règles de bonne conduite ne sont pas seulement des exigences du contrôle social, mais qu'elles sont constitutives d'avantages sociaux aux conséquences effectives¹⁴⁶». Disposer de capital symbolique, c'est donc avoir un certain pouvoir, une certaine influence, sur ceux qui sont prêts à nous accorder du crédit¹⁴⁷.

Une notion centrale du capital est sa possibilité de conversion sous une autre forme¹⁴⁸. Ces conversions génèrent des coûts ou des profits, dont les taux de conversion sont relatifs aux capitaux impliqués et au champ dans lequel l'agent s'insère¹⁴⁹. Les agents visent ainsi à augmenter leur quantité de capitaux pour obtenir une meilleure position dans l'espace social. Pierre Bourdieu illustre ses théories de l'espace et des champs sociaux à l'aide d'un diagramme représentant, dans l'espace social, la répartition des champs et des capitaux sociaux échangés (figure 10).

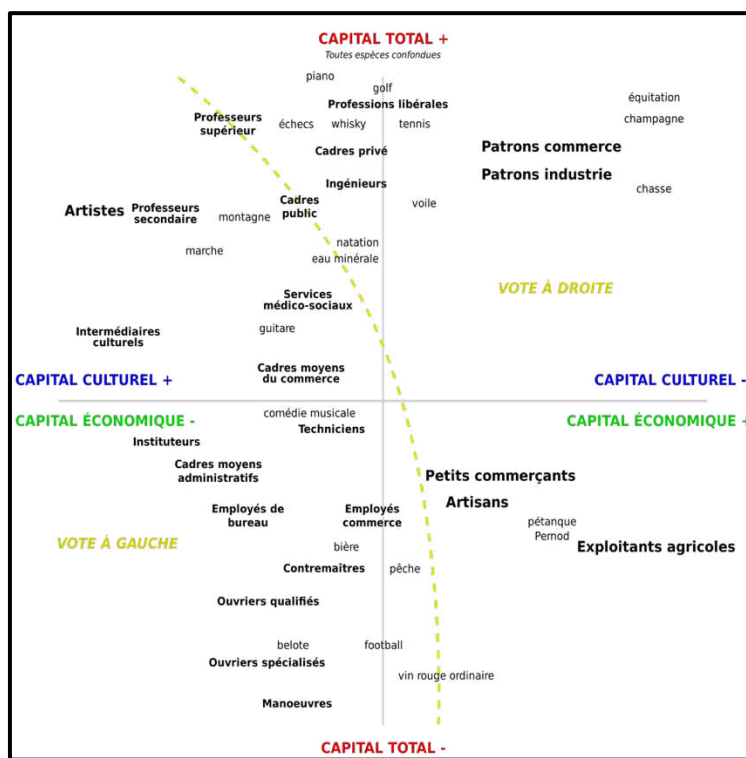


Figure 10 L'espace social selon Pierre Bourdieu¹⁵⁰

Pour illustrer sa conceptualisation, Pierre Bourdieu¹⁵¹ propose l'analogie suivante : il compare l'espace social à un jeu de cartes, dans lequel les acteurs sont assimilables aux joueurs, et les capitaux sociaux comparables aux cartes.

¹⁴⁶ Patrice Bonnewitz, *Pierre Bourdieu : vie, œuvres, concepts*, Paris, Ellipses, 2002, 94 p., p. 43.

¹⁴⁷ Anne Jourdain et Sidonie Naulin, *La sociologie de Pierre Bourdieu, op. cit.*

¹⁴⁸ Pierre Bourdieu, « The forms of Capitals », *op. cit.*

¹⁴⁹ *Ibid.*

¹⁵⁰ Source : Nicolas Lardot, schéma simplifié et inspiré du schéma de Pierre Bourdieu dans Pierre Bourdieu, *Raisons pratiques*, Seuil, Paris, 1996, 251 p. Œuvre publiée sous la licence GNU Free Documentation Licence version 1.2 ou ultérieure, libre de partage.

¹⁵¹ Pierre Bourdieu, « Espace social et genèse des "classes" », *op. cit.*

De la même manière que dans un jeu, les joueurs échangent des cartes – plus ou moins bonnes – selon les règles, dans un champ social, les acteurs échangent des capitaux sociaux – plus ou moins valorisants – selon les règles sociales spécifiques au champ dans lequel ils évoluent. Le but du jeu n'est pas nécessairement d'amasser le plus de cartes, mais plutôt de trouver des stratégies pour maximiser la valeur des cartes à disposition, voire de se procurer les meilleures cartes. Ces stratégies seront influencées par les cartes possédées ainsi que par le cadre évolutif du jeu. En fonction du champ social dans lequel les acteurs évoluent, les quatre formes de capital précédemment décrites (et leurs combinaisons) vont être plus ou moins avantageuses pour les acteurs et leur permettre ou non d'évoluer dans le champ.

Pour mieux appréhender les structures sociales formées par les réseaux criminels, nous nous intéresserons dans la suite de ce chapitre aux théories issues de la sociologie de la déviance, qui reviennent sur le processus de criminalisation des individus.

b) Développement de la sociologie de la déviance

Historiquement, la sociologie de la déviance s'est constituée à partir de la question de la délinquance au milieu du XIX^{ème} siècle. La déviance est la transgression d'une norme. Elle implique l'adoption d'un type de comportement, et l'existence d'une norme le prohibant. Les normes sont un ensemble d'incitations, d'obligations ou d'interdictions fondées sur un socle de valeurs visant à orienter le comportement des individus en société. Elles peuvent être formelles (lois, règlements, codes...) ou informelles (usages, coutumes, traditions, morale...).

La sociologie de la déviance vise autant à comprendre le passage à l'acte, par le biais de l'étude des causes, qu'à définir la déviance en tant que processus, c'est-à-dire en explicitant les mécanismes¹⁵². Selon Xavier de Larminat, trois grands types de théories permettent d'expliquer le passage à l'acte (figure 11) : le déterminisme, l'interactionnisme et l'individualisme.

¹⁵² Xavier De Larmina, *Sociologie de la déviance : des théories du passage à l'acte à la déviance comme processus*, <http://ses.ens-lyon.fr/articles/sociologie-de-la-deviance>, consulté le 6 mai 2020 ; Albert Ogien, *Sociologie de la déviance*, 2018, 261 p.

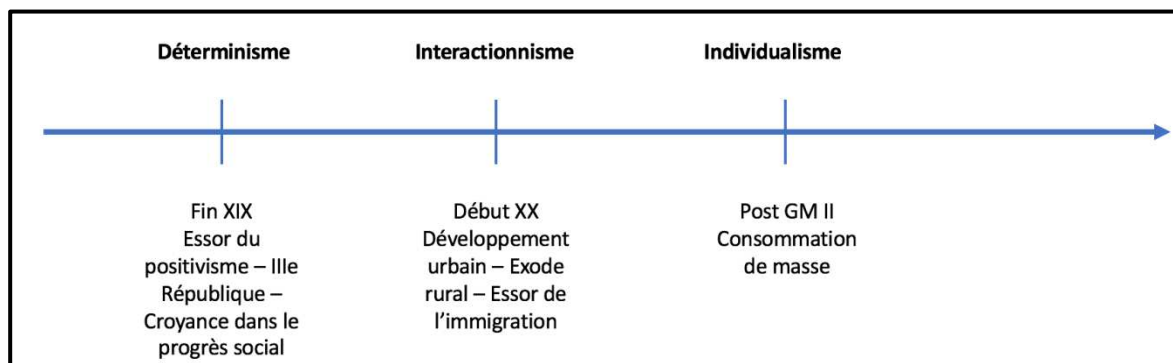


Figure 11 Apparition des trois grandes théories du passage à l'acte

Si ces courants sont apparus successivement, rien n'indique pour autant qu'un modèle en ait remplacé un autre. Ainsi, pour avoir une vision plus globale et plus exhaustive, il est important de considérer l'hybridation de ces modèles entre eux, *i.e.* de les voir comme complémentaires les uns des autres, comme l'illustrent les développements suivants.

1. Le déterminisme

Le déterminisme est un courant de pensée né à la fin du XIX^{ème} siècle qui stipule que la déviance provient de fondements biologiques, génétiques, neurologiques ou d'une pathologie affectant certains types d'individus (atavisme biologique lié à l'individu). L'idée en est qu'un individu socialement déviant présentait nécessairement des anomalies. En particulier, le déterminisme biologique pour expliquer les causes de la déviance est une théorie d'anthropologie criminelle exposée notamment en 1876 par Cesare Lombroso, professeur de psychiatrie italien. Dans ces premiers travaux, il a cherché à quantifier le poids de l'atavisme en conduisant des observations détaillées de plusieurs groupes de reclus liés à des enquêtes de même nature menées en Europe et aux États-Unis. En comparant les données morphologiques et statistiques recueillies, il conclut que le délinquant serait une « espèce particulière d'être humain, aisément reconnaissable à certains attributs physiques héréditairement transmis », tels que notamment un système pileux peu développé, le front fuyant, ou encore une mâchoire et des zygomés extrêmement développés¹⁵³. Ainsi, selon ce courant, le comportement déviant d'un individu est inné, inscrit dès sa naissance dans son génome, et selon Cesare Lombroso, ces gènes donnent lieu à des traits morphologiques identifiables. Ces postulats ont été réfutés dès le début du XX^{ème} siècle, notamment avec le développement de l'École de Chicago et du courant interactionniste, et, plus tard, de l'individualisme, qui postulent que la déviance relève d'un comportement acquis.

¹⁵³ Albert Ogien, *Sociologie de la déviance*, *op. cit.*, p. 23.

2. L'interactionnisme (École de Chicago)

Émile Durkheim conclut à la fin du XIX^{ème} que le crime n'est pas un phénomène pathologique, mais bien un phénomène normal, dans la mesure où il existe dans toutes les sociétés, quelles qu'elles soient, et quelles que soient leurs normes. Il est en cela l'un des premiers à envisager le crime comme la transgression d'une norme : « un acte est criminel quand il offense les états forts et définis de la conscience collective. [...] Nous ne réprouvons pas un acte parce qu'il est criminel, il est criminel parce que nous le réprouvons »¹⁵⁴.

L'interactionnisme est un courant de pensée apparu au début du XX^{ème} siècle, dans un contexte de développement urbain et d'essor de l'immigration. Il propose d'étudier les phénomènes de déviance en s'intéressant aux relations entre les individus ainsi qu'à leurs interactions avec leur environnement (atavisme social lié à la société).

Robert K. Merton met l'accent sur la structure sociale pour expliquer la déviance. Il distingue ainsi les buts poursuivis par les membres d'une société et les moyens mis en œuvre pour y arriver. Il expose que les voleurs, partageant la même aspiration économique que les citoyens ordinaires, ne s'en distinguent pas par leurs buts, mais bien par les moyens mis à leur disposition pour les atteindre. « La déviance apparaît alors comme une réponse à la contradiction entre les valeurs de la société et les conditions réelles d'existence qu'elle propose à sa population »¹⁵⁵. Pour rejoindre les conclusions d'Émile Durkheim, ce ne sont pas ici les buts que la société réproouve, mais les moyens employés. Cette approche semble très pertinente dans le cas des acteurs de trafics de stupéfiants : très souvent, aux plus bas niveaux de la chaîne¹⁵⁶, ces activités sont perçues comme une source d'argent facile, en opposition à l'inégalité des chances défavorable aux individus observée dans les milieux les plus concernés (cités, quartiers populaires...). Cependant, il serait trop restrictif d'attribuer les comportements criminels, en particulier dans le cadre de réseaux, au simple manque de moyens.

¹⁵⁴ Émile Durkheim, *De la division du travail social*, Paris, Quadrige, 1998, 416 p., p. 85-86.

¹⁵⁵ Robert K. Merton, « Social Structure and Anomie », *op. cit.*, p. 675.

¹⁵⁶ Le « trafic de stupéfiant » est vaste : il contient toute la chaîne depuis la production jusqu'au réinvestissement d'argent blanchi. Sans pouvoir à ce stade des recherches poser une affirmation éclairée et documentée, il est nécessaire d'évaluer les motivations et contraintes aux différents niveaux de la chaîne (bas de l'échelle, management du réseau, criminalité en col blanc : corruption/blanchiment).

Cette approche pourrait poser un des critères de la distinction entre les « membres » du trafic, et les « employés ». D'autres critères sont très certainement à prendre en compte (comme le sentiment d'appartenance à un groupe), mais pour les « employés » des trafics, cette activité constituerait plus un moyen d'atteindre les mêmes buts que les citoyens ordinaires, tandis qu'au contraire, pour les « membres » du réseau, le trafic en soi pourrait être un but.

En effet, selon William I. Thomas, les habitudes et l'éducation propres à une communauté donnée font acquérir ses valeurs à ses acteurs ainsi que le respect des normes de conduite qui y sont établies – implicitement ou explicitement¹⁵⁷.

La théorie de la désorganisation sociale est un point de départ de ce courant. William I. Thomas et Florian Znaniecki la définissent en 1919 comme suit : « la désorganisation sociale est la dissolution des contrôles sociaux en raison de la transition d'un mode de vie communautaire à des modes de vie plus urbains, et donc des modes de contrôle plus anonymes au sein des villes »¹⁵⁸. Pour ces auteurs, « la désorganisation [d'une] communauté commence lorsque ses membres [...] dissocient l'opinion de leurs pairs à propos d'une chose de l'importance qu'ils lui accordent »¹⁵⁹. Les zones rurales fonctionnent sur la base de « contrôles primaires » directs (familles, voisinages...) du respect des individus des normes établies par la communauté, et la désorganisation intervient lorsque certains membres – de par l'urbanisation et la mixité induite par l'immigration – se retrouvent à la frontière entre deux systèmes de valeurs communautaires concurrents. Les zones urbaines, qui sont par construction une concentration de communautés, fonctionnent en outre sur la base de « contrôles secondaires » anonymes et impersonnels (par le biais de diverses institutions, comme l'église, ou de la communauté au sens large : « les gens »). Le lien entre les individus étant plus relâché, les contrôles sont moins efficaces, moins dissuasifs¹⁶⁰.

Toujours dans le courant interactionniste, mais avec une autre approche, Frederic Thrasher, en 1927, cherche à situer les territoires où se développe la délinquance, et montre que la ségrégation urbaine n'est pas seulement sociale et ethnique, mais aussi spatiale. Il définit ainsi des zones où la délinquance est particulièrement développée : les zones interstitielles (entre les centres administratifs et commerciaux et les périphéries résidentielles). Selon lui, la localisation territoriale prime sur la couleur de peau pour expliquer le passage à l'acte délinquant (de nombreuses bandes sont mixtes plutôt qu'ethniquement homogènes)¹⁶¹. Ce concept d'analyse géographique et spatiale de la délinquance sera au XXème siècle intégré dans la « théorie des *patterns* criminels » émise lors du développement du modèle d'*intelligence-led*

¹⁵⁷ William Isaac Thomas, *On social organization and social personality*, Chicago, The University of Chicago Press, 1966, 311 p.

¹⁵⁸ William Isaac Thomas et Florian Znaniecki, *Le paysan polonais en Europe et en Amérique : récit de vie d'un migrant*, Paris, A. Colin, 2005, 446 p.

¹⁵⁹ William Isaac Thomas, *W.I. Thomas on social organization and social personality*, op. cit., p. 70.

¹⁶⁰ Robert Ezra Park, Ernest Watson Burgess et Roderick D. McKenzie, *The city: suggestions for investigation of human behavior in the urban environment*, Chicago, Univ. of Chicago Press, 2010, 239 p.

¹⁶¹ Frederic Milton Thrasher, *The gang: a study of 1,313 gangs in Chicago*, Chicago, Illinois, The University of Chicago Press, 2013, 605 p.

*policing*¹⁶². Jerry Ratcliffe proposera notamment une classification des lieux criminogènes en trois termes : les lieux générateurs de crime (augmentation des cibles), les lieux attracteurs de crime (augmentation des auteurs), et les lieux activateurs de crime (diminution des gardiens)¹⁶³.

En 1943, William F. Whyte expose que les déviations observées dans ces zones traduisent une forme de socialisation concurrente de la norme dominante : certains modèles, tels que celui du « gang », sont bel et bien une forme spécifique d'organisation sociale¹⁶⁴. Ainsi se développe la notion de sous-culture délinquante, véritable déclinaison (sous-ensemble) de la société dotée de ses propres normes et règles. Selon chaque sous-culture, différents capitaux seront valorisés, dépendant des motivations des individus qui la composent, des codes adoptés, mais aussi et surtout des raisons d'être et ambitions de ce sous-ensemble sociétal.

Cette notion de sous-culture délinquante intéresse également Albert K. Cohen, qui propose en 1955 une étude de la délinquance juvénile qui se construit sur trois principes¹⁶⁵ :

- i. La signification qu'accordent les délinquants aux actes déviants se manifeste dans leurs interactions ;
- ii. Selon l'auteur et sa manière de procéder à l'infraction, ainsi que les conceptions qu'en ont les acteurs d'un groupe social, la signification de l'infraction varie ;
- iii. Ces conceptions se construisent dans le cadre de sous-cultures données, qui se sont constituées en réponse aux contraintes spécifiques auxquelles ces groupes sociaux sont confrontés par l'ordre social dans lequel ils évoluent – tout en s'en dissociant.

À l'inverse des théories de la désorganisation sociale, la notion de culture ne renvoie pas ici au fait que les valeurs d'un groupe social soient issues d'une tradition et ainsi transmises comme support de l'identité du groupe, mais renvoie plutôt au fait qu'un ensemble d'individus se regroupe parce qu'ils partagent un monde de valeurs, forgeant leurs propres normes¹⁶⁶. Ainsi, pour Albert K. Cohen, la sous-culture délinquante propre aux bandes de jeunes présente trois caractéristiques : elle est non-utilitaire, dans la mesure où l'objectif de l'activité criminelle n'est pas instrumental ; elle est malintentionnée, l'activité criminelle relève du défi, de la nuisance ;

¹⁶² ILP : « Modèle organisationnel et une philosophie de management dans lequel l'analyse des données et le renseignement criminel jouent un rôle central dans la prise de décision relative à la réduction de la criminalité, et la prévention de la délinquance, grâce à la fois à une stratégie d'organisation, et une doctrine judiciaire pragmatique ciblant les délinquants sérieux et récidivistes » (Ratcliffe, 2016).

¹⁶³ Jerry Ratcliffe, *Intelligence-led policing, op. cit.*

¹⁶⁴ William Foote Whyte, *Street corner society: la structure sociale d'un quartier italo-américain*, Paris, La Découverte, 1943, 406 p.

¹⁶⁵ Albert K. Cohen, *Delinquent Boys: The Culture of the Gang*, New York, The Free Press, 1955, 178 p.

¹⁶⁶ Albert Ogien, *Sociologie de la déviance*, 2018, 261 p.

et négative, c'est-à-dire que l'activité criminelle s'oppose systématiquement aux normes sociétales établies principes¹⁶⁷.

Richard Cloward et Lloyd Ohlin affinent et discutent cette position en 1966 en distinguant trois types de sous-cultures délinquantes¹⁶⁸:

- i. Sous-culture criminelle : délinquance d'appropriation des biens, qui découlerait de la frustration liée à l'absence d'opportunités légitimes et de l'existence préalable de groupes délinquants structurés. L'hétérogénéité des âges favorise le recrutement, la formation et l'établissement d'une hiérarchie interne.
- ii. Sous-culture conflictuelle : usage de la violence et manifestations collectives de virilité, comportant une forte dimension initiatique et intégratrice. Logique de surenchère et d'auto-valorisation exacerbée sans autres modes de distinction (homogénéité des profils).
- iii. Sous-culture de retrait ou d'évasion : liée à la consommation de la drogue, l'entrée y serait le fruit de l'échec à faire sa place dans la sous-culture criminelle et/ou conflictuelle, et de l'absence d'autres voies légitimes de réussite.

Si elles permettent d'apporter une explication d'ensemble des déviations, l'un des problèmes des approches accordant une grande importance à l'influence de l'environnement socio-économique, est qu'elles permettent plus difficilement de comprendre les différences existant à l'intérieur d'un même groupe social¹⁶⁹.

Ainsi, si l'approche déterministe pose la criminalité (en tant que transgression des lois) comme une réponse par nécessité, l'interactionnisme la pose comme une volonté seule de se créer une identité en rupture avec la norme. À nouveau, il apparaît nécessaire de compléter ces approches, plutôt que de classer une situation générale dans l'une ou l'autre.

3. L'individualisme

L'individualisme pose l'individu en tant que tel au centre des études sociétales. Apparue en économie, politique et philosophie dès le XVIII^e siècle, cette approche a été reprise par les sociologues après-guerre, dans un contexte de consommation de masse, en réaction aux théories déterministes structuralistes¹⁷⁰. La sociologie, pour expliquer le social, prend en

¹⁶⁷ Albert K. Cohen, *Delinquent Boys: The Culture of the Gang*, op. cit.

¹⁶⁸ Richard Cloward et Lloyd Ohlin, *Delinquency and opportunity*, Free Press, 1966, 230 p.

¹⁶⁹ Xavier De Larmina, *Sociologie de la déviance : des théories du passage à l'acte à la déviance comme processus*, op. cit.

¹⁷⁰ Jean-Pierre Delas et Bruno Milly, « Les individualismes méthodologiques », dans *Histoire des pensées sociologiques*, Paris, Armand Colin, 2015, pp. 371-411.

compte les motivations intrinsèques de l'Homme à agir (théorie économique du choix rationnel, interprétation hédoniste). En effet, une manière de répondre aux objections soulevées par les deux premiers courants est de faire appel à des explications à l'échelle individuelle. L'individualisme méthodologique est une perspective qui considère que bien que le tout ait une énergie propre, le social peut être vu comme le résultat de la somme des actions individuelles des membres de la société.

Cette approche postule en outre que les acteurs agissent de façon rationnelle dans la plupart des cas (en fonction de leurs connaissances et interprétations des situations).

i. Analyse économique : théorie du choix rationnel (gain/coût)

Selon Albert Ogien, la théorie du choix rationnel peut être définie comme la « recherche du maximum de satisfaction et d'un minimum d'efforts pour atteindre un objectif déterminé »¹⁷¹. Cette théorie a été appliquée au comportement criminel par l'économiste Gary Becker dans son article intitulé « *Crime and Punishment* »¹⁷². Selon lui, les « coûts » d'une infraction se répartissent en deux catégories : ceux liés à l'infraction commise en soi (dépenses engendrées, atteintes physique, réputation), et ceux liés à sa répression (justice, emprisonnement, isolement, manque à gagner).

Cette perspective est cependant très controversée, en particulier au sein des sciences sociales. On peut en citer quelques raisons, dont la première est son incompatibilité avec la dimension morale et symbolique des infractions et sanctions : comment calculer le bénéfice d'un meurtre ou le coût de l'emprisonnement ? Deuxièmement, cette perspective ne prend pas en compte le fait que les acteurs ne disposent pas de toutes les données pour prendre leur décision, ils ne peuvent calculer précisément la probabilité qu'ils ont d'être arrêtés, jugés et condamnés, qui dépend de facteurs qu'ils ne maîtrisent jamais totalement : leur rationalité est donc limitée. Enfin, elle décontextualise la commission d'infractions, ne considère pas l'influence ou interférence extérieure, et oublie la théorie des opportunités. En particulier, la dimension collective des déviations ainsi que la part d'incertitude liée aux interactions sociales ne sont pas considérées.

¹⁷¹ Albert Ogien, *Sociologie de la déviance*, op. cit., p. 178.

¹⁷² Gary Becker, « Crime and Punishment: An Economic Approach », *Journal of Political Economy*, vol. 76, n° 2, 1968, pp. 169-217.

ii. Analyse stratégique (gain/effort/risque)

Maurice Cusson, quant à lui, s'éloigne du cadre purement économique et développe une perspective individualiste qu'il appelle « analyse stratégique », au sens où les délinquants agiraient comme des stratèges militaires sur le champ de bataille. Il cherche à identifier ce qui pousse les délinquants à agir, parlant ainsi de paradigme actionniste. Il n'inclut pas uniquement des paramètres de nature économique : sont aussi intégrées des variables plus subjectives, telles que le prestige ou la notion d'influence.

Il distingue ainsi quatre finalités du délit¹⁷³ : l'action, par opposition à l'ennui (caractère ludique ou festif de la transgression), l'appropriation (impératif de survie/logique consumériste), l'agression (vengeance/racisme/logique utilitaire) et la domination (désir de puissance/de cruauté). Quelle que soit la finalité, les risques existent pour l'auteur, qu'ils soient physiques (blessures, incarcération), moraux (perte de l'estime des proches, menace sur la vie affective) ou sociaux (perte d'emploi, de relations sociales). La théorie de l'analyse stratégique appréhende ainsi différemment de la théorie du choix rationnel les « coûts » d'une infraction, en distinguant les risques encourus des efforts à fournir lors de l'acte criminel (environnement, temps, argent, investissement personnel...).

Les questions posées ne concernent plus uniquement les causes de la déviance, mais aussi son procédé de construction : il s'agit alors d'expliquer « comment » un individu devient déviant. S'inscrivant dans la lignée des théories des sous-cultures délinquantes et de l'association différentielle d'Edwin Sutherland¹⁷⁴, le courant individualiste est un passage des théories causales qui cherchent à identifier des facteurs individuels, sociétaux et environnementaux du passage à l'acte, aux théories compréhensives de la déviance, étudiant les processus évolutifs que constitue le passage à l'acte.

Cette nouvelle vision repense la déviance comme « une propriété non du comportement lui-même, mais de l'interaction entre la personne qui commet l'acte et celles qui réagissent à cet acte¹⁷⁵ ». Dès lors, la déviance ne peut être considérée ni comme une propriété interne liée aux caractéristiques des individus, ni comme le sous-produit de l'environnement social. Elle serait en fait le résultat d'une action collective de désignation et de labellisation des comportements.

¹⁷³ Maurice Cusson, *Délinquants, pourquoi ?*, Chicoutimi, Bibliothèque québécoise, 1995, 295 p.

¹⁷⁴ Cf II.1 c) « Le crime comme occupation » et d) « Les capitaux dans l'espace social criminel ».

¹⁷⁵ Howard S. Becker, *Outsiders: Studies in the Sociology of Deviance*, Glencoe, 1963, 224 p., traduction de l'ouvrage de l'anglais vers le français par J.-P. Briand.

Dès 1938, Franck Tannenbaum suggérait que la délinquance ne trouve pas sa source dans la personne qui commet une infraction mais dans l'étiquette que lui a initialement collée la société pour des actes sans importances. Faute de se défaire de cette étiquette par la suite, l'individu devient un véritable délinquant¹⁷⁶.

Ainsi l'idée de carrière déviante commence à voir le jour. Everett Hughes, cité par Xavier De Larmina, définit ainsi la notion de carrière : « le parcours ou la progression d'une personne au cours de la vie – ou d'une partie donnée de celle-ci »¹⁷⁷.

L'un des principaux intérêts de la notion de carrière est d'articuler des dimensions objectives et subjectives : prise en compte des structures sociales, de la situation officielle d'un individu, ainsi que de sa propre manière de voir les choses et de se représenter sa situation. L'objectif est de considérer un comportement comme le résultat d'une succession de différentes séquences ordonnées les unes par rapport aux autres. Ces séquences désignent les étapes récurrentes dont tous ceux qui partagent une même carrière ont fait l'expérience. Cependant, la manière dont ces séquences s'organisent, leur durée, les contextes et relations qui interviennent au cours de ce processus varient, d'où l'apparente singularité de chaque carrière.

À chaque séquence de la carrière, Howard S. Becker observe que, en fonction de l'enchaînement des relations et des événements, certains individus vont s'arrêter tandis que d'autres vont persévérer et progresser dans la carrière en accédant à la séquence suivante¹⁷⁸. Contrairement aux théories du passage à l'acte, pour lesquelles on cherchait à identifier des causes universelles et intangibles du passage à l'acte déviant, la subtilité de la notion de carrière réside dans le fait que chaque séquence requiert une explication qui lui est propre (un élément jouant un rôle capital à un moment de la carrière peut avoir ensuite une importance négligeable – et inversement).

Enfin, plus un individu avance dans sa carrière, plus les choix qui s'offrent à lui sont restreints par effet d'entonnoir : les décisions déjà prises tendent à limiter les alternatives qui restent ouvertes, comme l'avait postulé Tannenbaum dès 1938¹⁷⁹ (tableau 1).

¹⁷⁶ Frank Tannenbaum, *Crime and the Community*, Columbia, Columbia University Press, 1938, 112 p.

¹⁷⁷ Xavier De Larmina, « Sociologie de la déviance : des théories du passage à l'acte à la déviance comme processus », *op. cit.*

¹⁷⁸ Howard S. Becker, *Outsiders*, *op. cit.*

¹⁷⁹ Frank Tannenbaum, *Crime and the Community*, *op. cit.*

Tableau 1 Synthèse de la théorie de l'étiquetage de Becker¹⁸⁰

	Obéissant à la norme	Transgressant la norme
Perçu comme déviant	Accusé à tort	Pleinement déviant
Non perçu comme déviant	Conforme	Secrètement déviant

Selon Howard S. Becker, le fait d'être désigné comme déviant, à tort ou à raison, est un élément central de la progression dans la carrière. Symétriquement, plus un individu est engagé dans une carrière déviante, plus il prend le risque d'être repéré et catégorisé comme tel¹⁸¹. Ces deux dimensions s'alimentent donc mutuellement.

c) Le crime comme occupation

Howard S. Becker, lorsqu'il introduit la notion de carrière, s'intéresse à ce qu'il appelle « carrière déviante¹⁸² ». Il adopte un point de vue individuel, en termes de perceptions propres à l'individu, et cherche à comprendre comment ce dernier devient déviant, c'est-à-dire transgresse une ou plusieurs normes¹⁸³. En 1986, Alfred Blumstein, introduit quant à lui le concept de « carrière criminelle¹⁸⁴ », qui s'intéresse plus aux événements du parcours de l'individu, constitutifs de la trajectoire criminelle à proprement parler, qu'à la perception de ces événements par l'individu lui-même. L'étude des « carrières criminelles » se fait par le biais d'outils plus quantitatifs, on peut presque parler de mathématisation des événements. Cette notion implique que le crime peut être vu dans une certaine mesure comme une occupation¹⁸⁵. Elle peut ainsi être comprise comme un outil permettant d'étudier la trajectoire criminelle – ou « CV criminel » d'un individu.

Avant la seconde guerre mondiale, la sociologie du travail s'intéressait à des catégories bien précises, et très restreintes : les emplois industriels et les métiers certifiés par l'université. L'interactionnisme amène une ouverture d'esprit quant à la compréhension des occupations au sens large¹⁸⁶. Si dans un premier temps les sociologues s'intéressent aux occupations licites, les criminologues, dans la lignée de Howard S. Becker, Everett Hughes et Alfred Blumstein et du concept de « carrière criminelle », reprennent rapidement ces notions, et étudient dans quelle

¹⁸⁰ Source : Xavier De Larmina <http://ses.ens-lyon.fr/articles/sociologie-de-la-deviance>.

¹⁸¹ Howard S. Becker, *Outsiders*, *op. cit.*

¹⁸² Il est par ailleurs à noter que la notion de déviance est plus vaste que celle de crime, ce dernier correspondant à la transgression d'une norme légale uniquement.

¹⁸³ Howard S. Becker, *Outsiders*, *op. cit.*

¹⁸⁴ Alfred Blumstein, *Criminal careers and « career criminals »*, Washington, DC, National Acad. Press, 1986, 458 p.

¹⁸⁵ Yanick Charette, *La perception du prestige des occupations illicites par des délinquants : une perspective sur les capitaux criminels*, Mémoire de maîtrise, Université de Montréal, 2010, 95 p.

¹⁸⁶ *Ibid.*

mesure elles sont ou non transposables aux occupations illicites. Notamment, Richard A. Cloward, Ned Polsky et Paul Letkemann posent les bases de ces raisonnements en mentionnant que des compétences (spécifiques ou non) sont nécessaires pour la réussite des occupations illicites comme pour les occupations licites¹⁸⁷. Ainsi, toute théorie des occupations illicites se trouve être un cas particulier des théories générales des occupations¹⁸⁸.

La théorie d'analyse stratégique (gain/risque/effort) apparaissant avec le courant individualiste est ici très pertinente. Le délinquant choisira la voie criminelle si, selon son calcul stratégique, celle-ci présente plus d'avantages que la voie légitime¹⁸⁹. Maurice Cusson expose que « les sanctions produisent leur effet intimidant quand elles conduisent un individu à renoncer à son projet criminel en le forçant à conclure que ses risques dépassent son espérance de gains »¹⁹⁰. La notion de sanction n'est pas ici employée uniquement pour désigner la peine judiciaire, mais aussi les éventuelles blessures, ainsi que la perte d'estime de l'entourage.

Clément Robitaille soulève un point intéressant en montrant que si la recherche s'est longuement attardée sur les coûts (risques et efforts), elle est demeurée silencieuse quant au niveau espéré de profit que peut rapporter la criminalité à ses auteurs¹⁹¹. Au-delà du revenu, le délinquant plein d'aspiration cherchera des occupations qui lui procureront d'autres bénéfices suffisamment intéressants pour surpasser les risques liés à celle-ci. Ainsi, en faisant le parallèle avec la sociologie du travail, le prestige de son occupation criminelle renvoie à un signe de la performance ou de la réussite criminelle, donc d'une forme de bénéfice intéressant aux yeux du délinquant, qui pèsera dans sa décision à poursuivre ou non sa « carrière criminelle »¹⁹².

En 1947, Edwin H. Sutherland émet l'une des théories majeures en criminologie : l'association différentielle. Selon cette théorie, le comportement criminel est appris par des interactions dans un processus de communication dans le cadre de relations personnelles privilégiées. L'apprentissage se fait par l'initiation à des méthodes techniques, mais surtout par

¹⁸⁷ Paul Letkemann, *Crime as work*, Englewood Cliffs, 1973, 182 p. ; Ned Polsky, *Hustlers, beats and others*, Chicago, 1967, 218 p. ; Richard A. Cloward, « Illegitimate Means, Anomie, and Deviant Behavior », *American Sociological Review*, vol. 24, n° 2, 1959, pp. 164-176.

¹⁸⁸ Ned Polsky, *Hustlers, beats and others*, op. cit.

¹⁸⁹ Yanick Charette, *La perception du prestige des occupations illicites par des délinquants : une perspective sur les capitaux criminels*, op. cit.

¹⁹⁰ Maurice Cusson, *La délinquance, une vie choisie : entre plaisir et crime*, Chicoutimi, Bibliothèque québécoise, 2008, 34 p., p. 86.

¹⁹¹ Clément Robitaille, « À qui profite le crime ? Les facteurs individuels de la réussite criminelle1 », *Criminologie*, vol. 37, n° 2, 2005, pp. 33-62.

¹⁹² Yanick Charette, *La perception du prestige des occupations illicites par des délinquants : une perspective sur les capitaux criminels*, op. cit.

l'initiation à des motifs, des rationalisations et des attitudes.¹⁹³ Ainsi, c'est au travers d'interactions avec d'autres agents sociaux de l'espace – voire du champ – criminel que les acteurs vont développer une grande part de leur « capital culturel criminel », c'est-à-dire les acquis et ressources utiles au développement de leur carrière criminelle.

Pierre Tremblay réitère et met à jour la théorie de Edwin H. Sutherland en mentionnant trois normes de conduite définissant le comportement d'un délinquant accompli : la performance, l'endurance et le consensus¹⁹⁴.

Par ce critère de consensus, Pierre Tremblay entend un critère d'engagement partisan manifestant « un préjugé favorable à l'endroit des compagnons de route qui ont opté pour la délinquance, particulièrement ceux qui essaient de se conformer à l'idéal-type du délinquant. En essayant de s'y conformer, on obtient en retour le respect des autres délinquants et celui, plus officieux, de ses adversaires ¹⁹⁵». Le concept de consensus délinquant donne « *également une cohérence et un sens aux activités illicites* ¹⁹⁶» et rend possible une conceptualisation du prestige des occupations illicites. Le prestige existe par cet accord sur ce qu'est une réussite et un échec, tant dans le milieu criminel que dans le milieu licite. Certains auteurs, tels que Donald J. Treiman¹⁹⁷, Francis T. Cullen et Bruce G. Link ¹⁹⁸ ont mené des études en demandant à des personnes « ordinaires » quelle était leur évaluation du prestige de diverses activités criminelles. Si la tendance est d'accorder peu de prestige à ces activités, une hiérarchie existe. Il est par ailleurs vraisemblable qu'une même étude réalisée sur des individus ayant eux-mêmes des activités illicites aurait des conclusions différentes.

d) Les capitaux dans l'espace social criminel

1. Le capital économique

Certains chercheurs, tels que Michael R. Gottfredson et Travis Hirschi¹⁹⁹, Jack Katz²⁰⁰, James Q. Wilson et Allan Abrahamse²⁰¹ soutiennent que le crime ne rapporte pas de gain

¹⁹³ Edwin H. Sutherland, *Principles of criminology*, New York, General Hall Inc., U.S., 1947, 704 p.

¹⁹⁴ Pierre Tremblay, *Le délinquant idéal : performance, discipline, solidarité*, Montréal, QC, Liber, 2010, 263 p.

¹⁹⁵ *Ibid.*, p. 55.

¹⁹⁶ *Ibid.*, p. 57.

¹⁹⁷ Donald J. Treiman, *Occupational prestige in comparative perspective*, New York, Academic Press, 1977, 536 p.

¹⁹⁸ Francis T. Cullen et Bruce G. Link, « Crime as an occupation. A New Look », *Criminology*, vol. 18, n° 3, 1980, pp. 399-410.

¹⁹⁹ Michael R. Gottfredson et Travis Hirschi, *A general theory of crime*, Stanford, Calif, Stanford University Press, 1990, 297 p.

²⁰⁰ Jack Katz, *Seductions of crime: moral and sensual attractions in doing evil*, New York, Basic Books, 1988, 367 p.

²⁰¹ James Q. Wilson et Allan Abrahamse, « Does crime pay? », *Justice Quarterly*, vol. 9, n° 3, 1992, pp. 359-377.

pécuniaire suffisamment intéressant pour qu'il puisse représenter un intérêt réel pour les criminels. Pourtant, selon d'autres, notamment Malin Åkerström²⁰², Richard A. Cloward²⁰³, Irving Piliavin²⁰⁴, Mercer L. Sullivan²⁰⁵ et Pierre Tremblay²⁰⁶, l'argent serait le moteur principal d'une « carrière criminelle » réussie. En effet, plusieurs études²⁰⁷ montrent que les activités criminelles peuvent rapporter des gains considérablement intéressants et non négligeables. En revanche, les revenus soutirés des occupations illicites semblent n'être intéressants que pour une petite partie de la population criminelle²⁰⁸. En outre, les criminels ayant effectivement des revenus profitables sont conscients que cela peut être de courte durée : leur activité peut cesser à tout moment. Ainsi, pour ceux ayant les plus gros revenus, une part est souvent mise de côté, que ce soit pour épargner, ou pour s'assurer une situation financière en centre de détention et à la sortie, en cas d'arrestation²⁰⁹.

2. Le capital culturel

Le capital culturel se réfère aux compétences et connaissances transmises, institutionnalisées ou non. Paul Letkemann s'est particulièrement intéressé aux compétences requises pour pratiquer des activités criminelles. Il distingue trois types de compétences qui peuvent être utiles à la réalisation d'un délit²¹⁰, soit les compétences mécaniques (capacités techniques, physiques, intellectuelles²¹¹), organisationnelles (capacités individuelles de leadership face au groupe, de planification dans l'exécution et la coordination du projet délictueux²¹²) et sociales (telles que « la gestion des victimes ²¹³» et le « contrôle de la tension »). Le délinquant doit faire l'acquisition progressive de ces trois types de compétences,

²⁰² Malin Åkerström, *Crooks and squares: lifestyles of thieves and addicts in comparison to conventional people*, New Brunswick, N.J., Transaction Books, 1995, 265 p.

²⁰³ Richard A. Cloward, « Illegitimate Means, Anomie, and Deviant Behavior », *op. cit.*

²⁰⁴ Irving Piliavin, Rosemary Gartner, Craig Thornton et Ross L. Matsueda, « Crime, Deterrence, and Rational Choice », *American Sociological Review*, vol. 51, n° 1, 1986, pp. 101.

²⁰⁵ Mercer L. Sullivan, « *Getting paid* »: *youth crime and work in the inner city*, Ithaca, Cornell University Press, 1989, 275 p.

²⁰⁶ Pierre Tremblay, « Attrition, récidive et adaptation », *Revue internationale de criminologie et de police technique et scientifique*, vol. 52, n° 2, 1999, pp. 163-178.

²⁰⁷ Bill McCarthy et John Hagan, « When Crime Pays: Capital, Competence, and Criminal Success », *Social Forces*, vol. 79, n° 3, 2001, pp. 1035-1060 ; Clément Robitaille, « À qui profite le crime ? », *op. cit.* ; Carlo Morselli et Pierre Tremblay, « Délinquance, performance et capital social : une théorie sociologique des carrières criminelles », *Criminologie*, vol. 37, n° 2, 2005, pp. 89-122.

²⁰⁸ Steven D. Levitt et Sudhir Alladi Venkatesh, « An Economic Analysis of a Drug-Selling Gang's Finances* », *Quarterly Journal of Economics*, vol. 115, n° 3, 2000, pp. 755-789.

²⁰⁹ Karim Mokhtari est auteur de l'ouvrage Karim Mokhtari et Charlie Carle, *Rédemption : itinéraire d'un enfant cassé*, Paris, Scrineo, 2013, 448 p. Entretien réalisé le 28/06/21 à Brest.

²¹⁰ Paul Letkemann, *Crime as work*, *op. cit.*

²¹¹ Telles que voler une voiture en moins de trente secondes par exemple.

²¹² Comme par exemple repérer des cibles ou victimes faciles.

²¹³ Gagner la confiance d'une victime en vue de l'escroquer par exemple.

bien que ceux-ci s'articulent différemment en fonction à la fois du champ criminel dans lequel il agit, de la fonction et de la position qu'il occupe.

3. Le capital social

Le capital social a une place essentielle dans la délinquance de par le côté illégal de celle-ci. En effet, la criminalisation d'un acte rend difficile sa réalisation par les moyens de contrôle mis en place pour le prévenir : la présence de contacts dans différentes sphères du processus est nécessaire à son bon déroulement afin de contourner les moyens préventifs existants²¹⁴. La présence d'un réseau social important et bien structuré aura donc un impact sur la performance dans le crime²¹⁵.

Certains acteurs peuvent avoir un rôle plus important au sein du réseau social d'un individu. Edwin H. Sutherland amène une réflexion intéressante à ce sujet : dans sa théorie des associations différentielles, le comportement criminel est acquis au travers de relations personnelles privilégiées²¹⁶. Carlo Morselli, Pierre Tremblay et Bill McCarthy testent l'hypothèse selon laquelle la présence d'un mentor amènera des avantages dans sa réussite criminelle de la même façon que lors d'occupations licites²¹⁷. Les auteurs en arrivent à la conclusion que la présence d'un mentor est un atout important, non seulement pour la réussite criminelle (connaissances/compétences, capital culturel), mais aussi dans la constitution d'un bon réseau de pairs délinquants (relations, capital social). Ces conclusions rejoignent les propos de Burt selon lesquels, pour certaines occupations dont l'accès est restreint, cela force les aspirants à utiliser un capital social « emprunté »²¹⁸. Les mentors deviennent ainsi des « prêteurs de capital social » pour leur protégé, permettant à leurs élèves d'augmenter leur capital²¹⁹.

4. Le capital symbolique

Cette forme de capital est souvent mise en parallèle avec la notion de sous-culture délinquante²²⁰. Ainsi, il s'agit toujours d'un prestige relatif au champ criminel dans lequel le

²¹⁴ Carlo Morselli et Julie Roy, « Brokerage qualifications in ringing operations », *Criminology*, vol. 46, n° 1, 2008, pp. 71-98.

²¹⁵ Carlo Morselli et Pierre Tremblay, « Délinquance, performance et capital social », *op. cit.*

²¹⁶ Edwin H. Sutherland, *Le voleur professionnel : d'après le récit d'un voleur de profession*, 1963, 162 p.

²¹⁷ Carlo Morselli, Pierre Tremblay et Bill McCarthy, « Mentors and criminal achievement », *Criminology*, vol. 44, n° 1, 2006, pp. 17-43.

²¹⁸ Ronald S. Burt, *Structural holes*, *op. cit.*

²¹⁹ Yanick Charette, *La perception du prestige des occupations illicites par des délinquants : une perspective sur les capitaux criminels*, *op. cit.*

²²⁰ Richard Cloward et Lloyd Ohlin, *Delinquency and opportunity*, *op. cit.* ; Ross L. Matsueda, Rosemary Gartner, Irving Piliavin et Michael Polakowski, « The Prestige of Criminal and Conventional Occupations: A Subcultural Model of Criminal Activity », *American Sociological Review*, vol. 57, n° 6, 1992, pp. 752.

délinquant évolue. Afin d'écrire son ouvrage *In search of respect*²²¹, Philippe Bourgois réside durant cinq années dans un ghetto d'East Harlem à New York, son objectif étant d'étudier le phénomène du crack. Étant donné les ressources socioéconomiques et culturelles insuffisantes, l'auteur montre comment la participation au crime devient une façon efficace, pour les habitants du secteur, d'obtenir le respect de leurs pairs. Ainsi, la position des trafiquants n'est peut-être pas à envier des individus de l'extérieur du quartier, mais l'est bel et bien pour ceux y vivant. Cette étude aide à mieux comprendre pourquoi une mesure de prestige permet d'analyser l'évolution d'une carrière dans le milieu criminel.

II.2 Principes d'analyse de réseaux sociaux

a) Analyse de réseaux - généralités

En avant-propos de cette section, il est important de définir les notions de réseau social et d'analyse de réseaux sociaux telles que nous les comprenons. Ainsi, un réseau social est « un ensemble d'entités reliées ou non entre elles²²² ». L'analyse de réseaux sociaux (ARS) – ou *Social Network Analysis (SNA)* – rassemble les outils théoriques et méthodologiques permettant l'analyse de données relationnelles, plutôt que celles d'attributs d'entités²²³.

Claire Bidart²²⁴ explicite deux manières de s'intéresser aux réseaux sociaux : les réseaux personnels et les réseaux complets²²⁵. Les premiers sont centrés sur un individu, dont on représente tout ou partie du réseau social. Pour des raisons de lisibilité, l'individu en question est souvent absent de la représentation, il s'agit de dessiner les liens que peuvent avoir entre elles les différentes entités de son réseau. Elle ne sera représentée que si l'on intègre des éléments tels que le sens, la fréquence, ou l'intensité des liens qu'elle entretient avec les autres

²²¹ Philippe Bourgois, *In search of respect: selling crack in El Barrio*, Cambridge, Cambridge University Press, 2014, 434 p.

²²² Rémi Boivin et Carlo Morselli, *Les réseaux criminels*, op. cit.

²²³ *Ibid.*

²²⁴ Claire Bidart est une sociologue française, membre du réseau thématique n°26 "Réseaux sociaux" de l'Association Française de Sociologie du groupe de recherche « Analyse de réseaux en Sciences humaines et sociales », du Conseil d'animation scientifique de l'Institut Méditerranéen de Recherche avancée (IMéRA), du comité de rédaction de la revue *Sociologie*.

²²⁵ Dans sa conférence enregistrée à l'IméRA le 17/11/2017 intitulée « Brève initiation à l'analyse des réseaux à partir de la sociologie », dans le cadre du séminaire doctoral interdisciplinaire du collège doctoral AMU dans le cadre du cycle « Analyse des réseaux et applications : approche interdisciplinaire ».

Claire Bidart, « Brève initiation à l'analyse des réseaux à partir de la sociologie » par Claire Bidart | IMéRA - Institut d'études avancées, <https://imera.univ-amu.fr/fr/media/breve-initiation-a-lanalyse-reseaux-a-partir-sociologie-claire-bidart>, consulté le 13 novembre 2022.

entités. Les seconds concernent l'ensemble des entités appartenant à un réseau donné, et les liens qui les unissent ou non. L'une des problématiques de cette approche est de déterminer où placer la limite des entités ou individus appartenant effectivement au réseau. Cette limite va en premier lieu dépendre des objectifs du chercheur, et des questions auxquelles l'analyse a vocation à répondre.

La littérature distingue deux grands types de représentation des réseaux sociaux²²⁶: les réseaux simples (ou *one-mode*) d'individus, où les protagonistes sont directement liés et les réseaux à deux niveaux (ou *two-modes*) où des acteurs vont être liés soit directement soit par le biais d'autres objets (par exemple : deux individus fréquentant un même endroit, utilisant un même véhicule...). L'ARS permet d'exploiter les informations selon trois approches : les entités (ou nœuds), les relations (ou liens) ou les réseaux dans leur ensemble (ensemble nœuds et liens). Cette dernière démarche, qui sera la nôtre, s'effectue au travers de la théorie des graphes.

b) Contributions de la théorie des graphes en ARS

Nous adopterons ici la définition de Stanley Wasserman et Katherine Faust, qui posent un graphe comme une représentation de relations non-dirigées entre des entités²²⁷. Cela signifie qu'un graphe est composé de deux ensembles d'objets : les nœuds (ou sommets), qui représentent les entités considérées, et les arêtes (ou liens), qui illustrent les relations existantes entre ces dernières.

Depuis le début du XXème siècle et de plus en plus, la théorie des graphes est utilisée et appliquée en ARS. Les principaux apports en étant les suivants : tout d'abord, le vocabulaire emprunté permet d'identifier et de désigner de nombreuses propriétés structurelles²²⁸ ; en deuxième lieu, cette théorie propose des solutions mathématiques pour quantifier et mesurer ces propriétés²²⁹ ; enfin, étant donné ces deux premiers outils que sont le vocabulaire et les mathématiques, la théorie des graphes permet de prouver des théorèmes sur les graphes, et par

²²⁶ Rémi Boivin et Carlo Morselli, *Les réseaux criminels*, op. cit. ; Stephen P. Borgatti, Martin G. Everett et Jeffrey C. Johnson, *Analyzing social networks*, op. cit.

²²⁷ Stanley Wasserman et Katherine Faust, *Social Network Analysis Methods and Applications*, op. cit.

²²⁸ Frank Harary et Robert Z. Norman, *Graph Theory as a Mathematical Model in Social Science*, Ann Arbor, University of Michigan, Institute for Social Research, 1953, 45 p.

²²⁹ Linton C. Freeman, « Turning a profit from mathematics: The case of social networks », *The Journal of Mathematical Sociology*, vol. 10, n° 3-4, 12/1984, pp. 343-360.

extension sur les représentations des structures sociales²³⁰. En effet, la théorie des graphes est un outil puissant non seulement pour démontrer une thèse, mais aussi pour tester une hypothèse.

1. Éléments de base en théorie des graphes : définitions et formalismes

Un graphe est dit « orienté » lorsque les arêtes entre paires de nœuds sont dirigées, c'est-à-dire qu'elles vont spécifiquement d'un nœud vers l'autre : on parle alors d'arcs entre les sommets, et de couples de nœuds. Lors d'une transaction entre deux individus, la relation est dirigée : l'argent circule de l'acheteur vers le vendeur (figure 12). L'ordre du graphe correspond au nombre de nœuds qu'il contient.

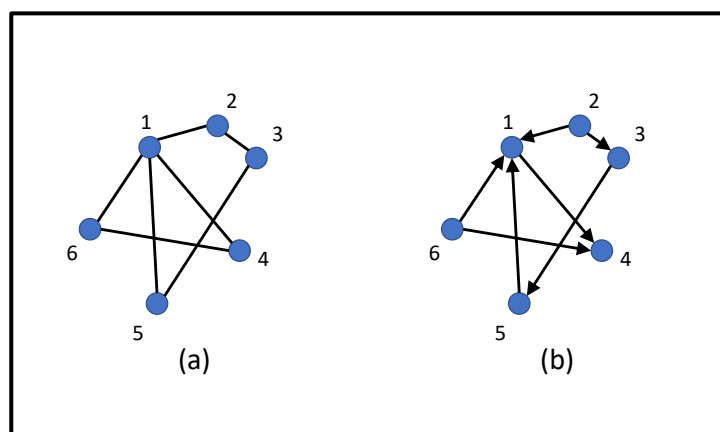


Figure 12 (a) Graphe G (non-orienté) d'ordre 6 ; (b) Graphe G' orienté d'ordre 6

Nous considérons dans la suite un graphe $G = (V, E)$ non orienté²³¹ d'ordre N avec $V = \{v_0, \dots, v_n\}$ l'ensemble des sommets et $E = \{e_0, \dots, e_m\}$ l'ensemble des arêtes. La matrice d'adjacence A de G recense les arêtes reliant les différents sommets. Il s'agit d'un tableau carré d'ordre $N \times N$, chaque ligne et chaque colonne correspondant à un nœud du graphe. Lorsque deux nœuds $(v_i, v_j) \in V$ sont reliés par une arête, l'entrée a_{ij} de la matrice correspondant à v_i en ligne et v_j en colonne est égale à 1, sinon elle vaut 0²³² (figure 13).

²³⁰ Frank Harary et Robert Z. Norman, *Graph Theory as a Mathematical Model in Social Science*, op. cit.

²³¹ Dans la mesure où nous traitons uniquement des graphes non-orientés dans le présent manuscrit, nous ne discuterons pas dans la suite des définitions, indicateurs et formalismes concernant les graphes orientés.

²³² Dans le cas d'un graphe non-orienté, la matrice d'adjacence est symétrique, i.e. $a_{ij} = a_{ji}$.

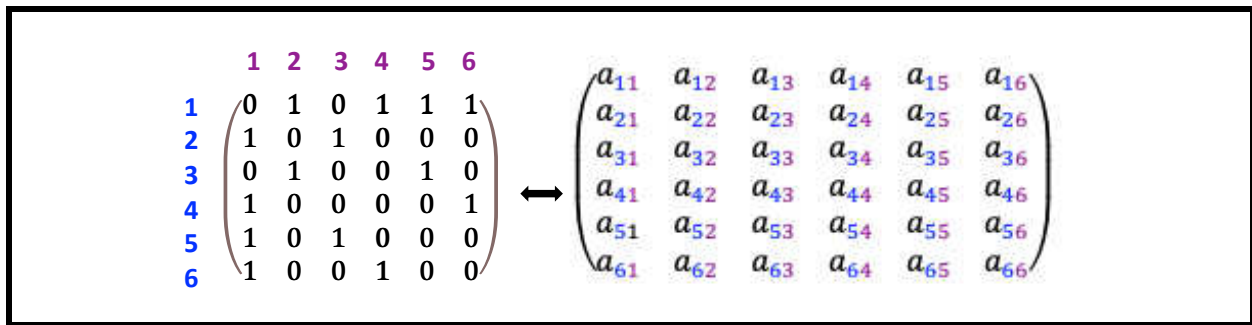


Figure 13 Matrice d'adjacence du graphe G en figure 12 (a)

Le vocabulaire en théorie des graphes²³³ désigne comme « chemin » une séquence de nœuds adjacents dans laquelle tous les sommets et toutes les arêtes sont distinct(e)s. La longueur d'un chemin est le nombre d'arêtes qu'il contient. Il peut exister plusieurs chemins, de différentes longueurs, entre une paire de nœuds $(v_i, v_j) \in V$ données. Les plus courts chemins, c'est-à-dire les chemins reliant v_i à v_j par le plus petit nombre de sommets intermédiaires, sont dits « géodésiques ». La distance entre deux nœuds $(v_i, v_j) \in V$ correspond à la longueur du (ou des) chemin(s) géodésique(s) entre v_i et v_j . Enfin, le diamètre d'un graphe est la longueur de son chemin géodésique maximal, c'est-à-dire la plus courte distance entre les deux nœuds du graphe les plus éloignés l'un de l'autre (figure 14).

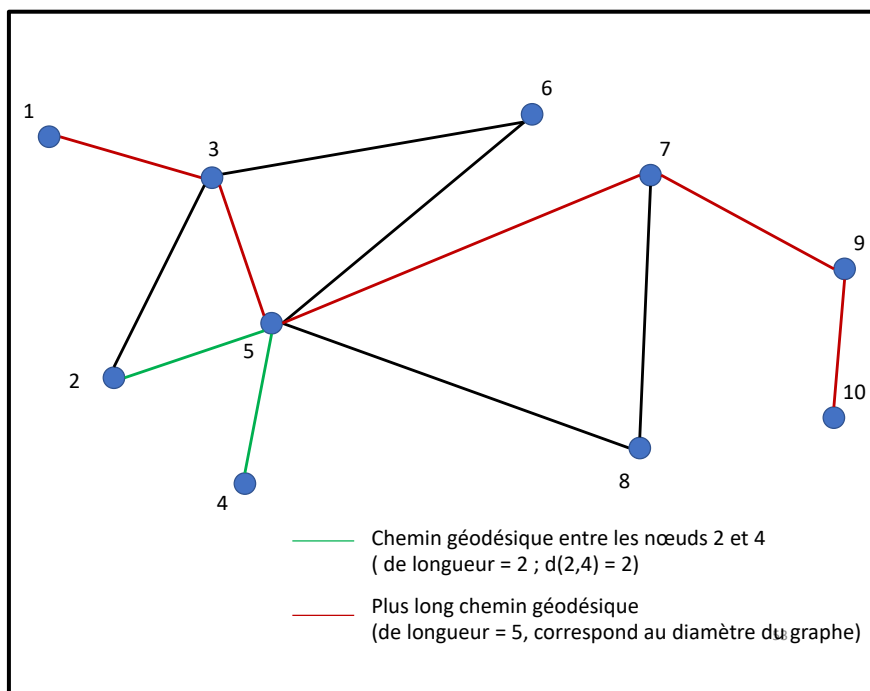


Figure 14 Illustration des définitions énoncées

²³³ Stanley Wasserman et Katherine Faust, *Social Network Analysis Methods and Applications*, op. cit.

2. Application de la théorie des graphes en ARS

Jacob Levy Moreno, médecin américain d'origine roumaine, est considéré dans la communauté des sciences sociales comme le père de la sociométrie, discipline précurseur de l'ARS. Ses recherches sur la compréhension des comportements sociaux et psychiatriques humains l'ont amené à inventer un moyen de représenter les structures interpersonnelles de groupe : le sociogramme²³⁴. Cette « invention », selon les propres termes de J. L. Moreno²³⁵, fût présentée pour la première fois en 1933 au cours d'un congrès universitaire de médecine. Un sociogramme est défini comme « une image dans laquelle des individus (ou, plus généralement, des unités sociales) sont représentés comme des points dans un espace bi-dimensionnel (les nœuds) et les relations entre paires de personnes par des lignes reliant les nœuds correspondants (les arêtes) »²³⁶.

Un sociogramme, ou graphe relationnel social, est un graphe particulier dans lequel les entités sont des personnes uniquement (éventuellement groupes de personnes), appelés « acteurs ». L'utilisation de la théorie des graphes en ARS permet d'obtenir une représentation d'un réseau social en tant que modèle de structure sociale, consistant en un ensemble d'acteurs et de liens les unissant. Stanley Wasserman et Katherine Faust caractérisent ainsi les sociogramme²³⁷ :

- i. Un sociogramme est un modèle relationnel représentant un réseau social avec des relations dichotomiques non dirigées. En d'autres termes, l'information contenue est, pour chaque paire d'acteurs du réseau, l'existence ou non d'un lien relationnel les unissant (adjacence ou non des acteurs), sans information de direction de l'un vers l'autre ou inversement. Il existe des graphes dans lesquels les relations entre les nœuds sont dirigées ; il s'agit alors de graphes orientés²³⁸.
- ii. La position des nœuds sur le schéma est arbitraire, et les longueurs des liens n'ont aucune signification : ils ont pour but de faciliter la lecture du sociogramme uniquement. Une fois encore, la seule information contenue est l'existence ou non de liens unissant les acteurs du réseau deux à deux.

²³⁴ J. L. Moreno et H. H. Jennings, « Statistics of Social Configurations », *Sociometry*, vol. 1, n° 3/4, 1938, pp. 342-374 ; Jacob Levy Moreno, « Sociogram and Sociomatrix: A Note to the Paper by Forsyth and Katz », *Sociometry*, vol. 9, n° 4, 1946, pp. 348-349.

²³⁵ Jacob Levy Moreno, *Who Shall Survive: A New Approach to the Problem of Human Interrelations*, Washington, DC, Nervous and Mental Disease Publishing Co., 1934, 460 p.

²³⁶ Stanley Wasserman et Katherine Faust, *Social Network Analysis Methods and Applications*, *op. cit.*, p. 12. Citation originale : « A sociogram is a picture in which people (or more generally social units) are represented as points in a two-dimensional space, and relationships among pairs of people are represented by lines linking the corresponding points ».

²³⁷ *Ibid.*

²³⁸ Cf. figure 12.

- iii. Dans le cas d'un sociogramme, la matrice d'adjacence est également appelée « *sociomatrix* ».

En outre, la représentation visuelle offerte par les sociogrammes permet de mettre en évidence des *patterns*, ou motifs répétés, qui ne sont pas détectables dans les données avec d'autres outils, tels que les matrices de données²³⁹. Les propriétés structurelles les plus classiquement étudiées sont les mesures de centralité, dont les plus usuelles et pertinentes dans notre cas sont décrites dans la sous-section suivante.

c) Mesures de centralité

Une première approche pour comprendre la structure sociale d'un réseau donné consiste à évaluer quels sont les acteurs sinon essentiels, déterminants dans le fonctionnement du réseau. Cela se fait notamment par la mesure puis la comparaison de la centralité de chacun des acteurs dans le réseau. La centralité d'un nœud dans un graphe renvoie à la place de ce nœud par rapport aux autres nœuds du graphe. Elle peut être calculée de plusieurs manières afin de répondre à des logiques et interprétations différentes. Ainsi, selon ce que l'on cherche à déterminer ou étudier, on va s'intéresser à l'une, à l'autre ou à l'ensemble de ces mesures.

1. La centralité de degré (*degree centrality*)

Le degré de centralité d'un nœud est calculé par simple comptage du nombre d'arêtes incidentes, rapporté au nombre de sommets du graphe²⁴⁰. Bien que ce nombre en soi donne peu d'information, il permet d'étudier l'importance relative des nœuds les uns par rapport aux autres dans un graphe. En effet, un acteur avec un fort degré de centralité a un potentiel d'interaction directe avec les autres plus important²⁴¹. On parle alors d'un acteur plus « actif » que les autres membres du réseau. Cependant, pour les acteurs d'intérêt, cette mesure mérite d'être complétée par d'autres mesures de centralité afin de cerner les niveaux d'influence et de maîtrise des flux de ces acteurs.

²³⁹ David C. Hoaglin, Frederick Mosteller et John Wilder Tukey (dir.), *Exploring data tables, trends, and shapes*, New York, Wiley, 1985, 527 p.

²⁴⁰ Alain Degenne et Michel Forsé, *Les réseaux sociaux*, op. cit.

²⁴¹ Paola Tubaro, *International encyclopedia of the social & behavioral sciences*, Amsterdam, Elsevier, 2015, 26 p ; Linton C. Freeman, « Centrality in social networks conceptual clarification », *Social Networks*, vol. 1, n° 3, 1978, pp. 215-239.

Soit $G = (V, E)$ un graphe non orienté d'ordre N représenté par sa matrice d'adjacence A^{242} . La centralité de degré $C_{degré}(v_i)$ du nœud $v_i \in V$ est définie comme suit²⁴³ :

$$C_{degré}(v_i) = \sum_j a_{ij}$$

2. La centralité de proximité (closeness centrality)

La centralité de proximité est une mesure de centralité globale, c'est-à-dire prenant en compte l'ensemble des sommets du graphe, à l'inverse de la centralité de degré qui ne tient compte que des sommets directement reliés au nœud v_i . Elle permet de déterminer si un nœud v_i est globalement proche de l'ensemble des autres nœuds du graphe et s'obtient en calculant sa proximité moyenne de vis-à-vis des autres nœuds²⁴⁴. En pratique, il s'agit de l'inverse de la somme des longueurs des chemins géodésiques entre le nœud d'intérêt v_i et chacun des autres nœuds, rapporté au nombre total de sommets du graphe²⁴⁵. Plus la valeur est grande pour un nœud donné, plus ce nœud a la possibilité d'accéder directement ou indirectement à un grand nombre d'autres nœuds *via* des trajets courts. Ainsi, une centralité de proximité élevée pour un acteur est un signe d'indépendance²⁴⁶ : il peut facilement recevoir ou passer de l'information, et accède rapidement aux communications et ressources qui passent par le réseau, sans avoir besoin de s'appuyer sur les autres membres. Pour les mêmes raisons, il peut aussi s'agir d'un signe d'influence.

²⁴² Voir figure 13.

²⁴³ Stanley Wasserman et Katherine Faust, *Social Network Analysis Methods and Applications*, *op. cit.*

²⁴⁴ Nacim Fateh Chikhi, *Calcul de centralité et identification de structures de communautés dans les graphes de documents*, Université Paul Sabatier - Toulouse III, Toulouse, 2010, 182 p.

²⁴⁵ Alain Degenne et Michel Forsé, *Les réseaux sociaux*, *op. cit.*

²⁴⁶ Michel Lacroix, « Littérature, analyse de réseaux et centralité : esquisse d'une théorisation du lien social concret en littérature », *Recherches sociographiques*, vol. 44, n° 3, 04/05/2004, pp. 475-497.

Soient $G = (V, E)$ un graphe non orienté d'ordre N et $d(v_i, v_j)$ la distance entre les sommets v_i et v_j . La centralité de proximité $C_{proximité}(v_i)$ du nœud $v_i \in V$ est définie comme suit²⁴⁷ :

$$C_{proximité}(v_i) = \frac{1}{\sum_j d(v_i, v_j)}$$

3. La centralité d'intermédierité (*betweenness centrality*)

La centralité d'intermédierité, qui est également une mesure globale, quantifie le niveau d'intermédierité d'un nœud v_i , c'est-à-dire le fait qu'il soit ou non un passage pour relier plusieurs nœuds du graphe²⁴⁸. Plus il existe de chemins géodésiques entre deux nœuds $(v_i, v_j) \in V$ du graphe passant par v_i , plus la centralité d'intermédierité de ce dernier est élevée. On parle alors de position stratégique – ou non²⁴⁹, et d'importance dans le réseau.

La centralité d'intermédierité identifie les « connecteurs » (ou « *brokers* ») dans le réseau, *i.e.* les acteurs permettant de relier *via* des trajets courts deux nœuds ou deux parties distinctes du réseau, parfois en étant un intermédiaire obligatoire²⁵⁰. D'avantage de chemins passent par les acteurs ayant une centralité d'intermédierité élevée, et donc d'avantage d'informations et de ressources, que ces acteurs peuvent ou non laisser passer. En plus de contrôler les flux d'information et/ou de ressources entre plusieurs parties du réseau, ces acteurs bénéficient d'un accès privilégié à de multiples sources, et peuvent agir en intermédiaires. Enfin, leur importance stratégique est non négligeable, dans la mesure où sans ces individus, certaines parties du réseau pourraient se trouver complètement déconnectées les unes des autres. La centralité d'intermédierité du nœud v_i correspond à la somme des rapports, pour chaque paire de nœuds $(v_j, v_k) \in V$, du nombre chemins géodésiques²⁵¹ entre v_j et v_k passant par v_i sur le nombre total de chemins géodésiques entre v_j et v_k .

²⁴⁷ Linton C. Freeman, « Centrality in social networks conceptual clarification », *op. cit.*

²⁴⁸ Michel Lacroix, « Littérature, analyse de réseaux et centralité », *op. cit.*

²⁴⁹ Carlo Morselli, « Assessing Vulnerable and Strategic Positions in a Criminal Network », *Journal of Contemporary Criminal Justice*, vol. 26, n° 4, 2010, pp. 382-392.

²⁵⁰ Stephen P. Borgatti, Martin G. Everett et Jeffrey C. Johnson, *Analyzing social networks*, *op. cit.*

²⁵¹ Un chemin géodésique est un plus court chemin entre deux nœuds, c'est-à-dire un chemin passant par un nombre minimal d'intermédiaires.

Soient $G = (V, E)$ un graphe d'ordre N , σ_{jk} le nombre total de chemins géodésiques entre les nœuds v_j et v_k et $\sigma_{jk}(v_i)$ le nombre de chemins géodésiques entre les nœuds v_j et v_k passant par v_i . La centralité d'intermédiarité $C_{intermédiarité}(v_i)$ du nœud $v_i \in V$ est définie comme suit²⁵² :

$$C_{intermédiarité}(v_i) = \sum_{j \in V} \sum_{k > j \in V} \frac{\sigma_{jk}(v_i)}{\sigma_{jk}}$$

4. La centralité de vecteur propre (eigenvector centrality)²⁵³

Une autre approche de la notion de centralité dans un graphe consiste à mesurer le niveau de connexion d'un nœud par rapport aux autres nœuds fortement connectés du graphe : la centralité de vecteur propre²⁵⁴. Cet indice est basé sur le principe qu'un lien avec un nœud très connecté « vaut » plus (*i.e.* procure plus de prestige, de pouvoir, d'information...) qu'un lien avec un nœud peu connecté²⁵⁵. En effet, les acteurs ayant une centralité de vecteur propre élevée peuvent facilement communiquer avec les acteurs les plus actifs (centralité de degré), influents (centralité de proximité) ou stratégiques (centralité d'intermédiarité). Enfin, cette mesure pondère les contacts sur la base de leur centralité, et tient compte de la structure du réseau dans son ensemble. Cet indicateur peut être vu comme une extension de la centralité de degré, dans la mesure où, au lieu de simplement compter les nœuds adjacents au nœud $v_i \in V$, on considère des poids attribués aux nœuds adjacents de v_i en fonction de leur propre centralité²⁵⁶. Phillip Bonacich propose de calculer la centralité de vecteur propre comme une combinaison linéaire des centralités de vecteur propre de ses nœuds voisins²⁵⁷.

²⁵² Linton C. Freeman, « Centrality in social networks conceptual clarification », *op. cit.*

²⁵³ Nacim Fateh Chikhi, *Calcul de centralité et identification de structures de communautés dans les graphes de documents*, *op. cit.*

²⁵⁴ Phillip Bonacich, « Some unique properties of eigenvector centrality », *Social Networks*, vol. 29, n° 4, 10/2007, pp. 555-564.

²⁵⁵ Matthieu Drevelle, « C comme centralité(s) ».

²⁵⁶ Nacim Fateh Chikhi, *Calcul de centralité et identification de structures de communautés dans les graphes de documents*, *op. cit.*

²⁵⁷ Phillip Bonacich, « Some unique properties of eigenvector centrality », *op. cit.*

Soient $G = (V, E)$ un graphe non-orienté d'ordre N représenté par sa matrice d'adjacence A^{258} , $v_i \in V$ le nœud d'intérêt, $(v_1, v_2, \dots, v_n) \in V$ les sommets qui lui sont adjacents, et $\mu \in \mathbb{R}^*$ un réel strictement positif²⁵⁹. La centralité de vecteur propre $C_{vect}(v_i)$ du nœud v_i est définie comme suit²⁶⁰ :

$$\mu \cdot C_{vect}(v_i) = a_{i1} \cdot C_{vect}(v_1) + a_{i2} \cdot C_{vect}(v_2) + \dots + a_{in} \cdot C_{vect}(v_n)$$

Le tableau 2 résume les interprétations conjointes des mesures de centralité de degré, proximité et intermédiarité.

Tableau 2 Interprétation de différentes mesures de centralité²⁶¹

	<i>Low Degree</i>	<i>Low Closeness</i>	<i>Low Betweenness</i>
<i>High Degree</i>		<i>Embedded in cluster that is far from the rest of the network.</i>	<i>Ego's connections are redundant: communication bypasses him/her.</i>
<i>High Closeness</i>	<i>Key player tied to important/active alters.</i>		<i>Probably multiple paths in the network, ego is near many people, but so are many people.</i>
<i>High Betweenness</i>	<i>Ego's few ties are crucial for network flow.</i>	<i>Very rare cell. Means that ego monopolizes the ties from a small number of people to many others.</i>	

²⁵⁸ Voir figure 13.

²⁵⁹ L'ensemble des réels désigne l'union des ensembles des nombres rationnels et irrationnels.

On écrit $\mathbb{R} = \mathbb{Q} \cup \mathbb{Q}'$. En pratique, un nombre réel est un nombre qui peut s'écrire sous la forme d'un entier suivi d'un nombre fini ou infini de décimales. On note \mathbb{R}^* l'ensemble des réels privé de 0, *i.e.* l'ensemble des réels strictement positifs.

²⁶⁰ Nacim Fateh Chikhi, *Calcul de centralité et identification de structures de communautés dans les graphes de documents*, *op. cit.*

²⁶¹ Source : Paola Tubaro, *International encyclopedia of the social & behavioral sciences*, *op. cit.*, adaptation d'une slide de James Moody.

II.3 Réseaux criminels : modélisations et connaissances

a) Visualisation relationnelle

On distingue deux approches générales, qui, loin d'être antinomiques, permettent une meilleure compréhension du système relationnel dans son ensemble lorsqu'elles sont combinées l'une avec l'autre. D'une part, l'approche visuelle, basée sur les entités, les relations et dans certains cas leurs attributs, qui a pour but de représenter les protagonistes, leurs rôles (fonctions), les relations, leur sens, leur intensité, leur force, ainsi que l'absence de relation entre certaines entités ; et d'autre part, l'approche sociométrique, fondée sur la combinaison de graphes et d'indicateurs, tels que les mesures de centralités, de densité, de détection de cliques. Il s'agit ici de comprendre la structure du système relationnel, de visualiser les positions (activité, pouvoir, influence) des acteurs et les atouts, en termes de capitaux sociaux que ces positions leur confèrent, ou réciproquement grâce auxquels les protagonistes accèdent à ses positions. Ces indicateurs facilitent notamment la détection d'acteurs et de relations clefs.

La complémentarité entre ces deux approches peut être résumée de la sorte : « On ne saurait se contenter de repérer l'existence de telle ou telle forme de “réseau” [approche visuelle], car il faut surtout se demander si ces “réseaux” sont ou non hiérarchisés, multipolaires, mouvants ou tenaces, à quels moments ils fonctionnent plutôt comme des ressources ou au contraire comme des contraintes pour les individus ou les groupes [approche sociométrique] »²⁶².

Ainsi, en procédant dans un premier temps à une visualisation des liens dans le réseau, puis à une analyse approfondie de la structure relationnelle, il devient possible de faire émerger les attributs, relations ou entités clefs, et ainsi d'orienter l'enquête et/ou le renseignement dans un but d'efficience et d'action proactive.

La conception d'un schéma repose sur la compréhension du problème (enquête) ou des objectifs de recherche (renseignement). En effet, il est rarement efficient (surcharge d'information rendant la lecture compliquée, voire impossible) de tout représenter sur un schéma relationnel. La première étape, après avoir déterminé les objectifs, consiste ainsi à identifier et classer les entités, relations et attributs pertinents pour permettre à l'analyste de

²⁶² Claire Lemercier, « Analyse de réseaux et histoire », *Revue d'histoire moderne & contemporaine*, vol. 52, n° 2, 2005, pp. 88-112., p. 88.

répondre à sa problématique. De fait, les schémas produits varient fortement suivant les objectifs d'analyse.

Quentin Rossy²⁶³ propose d'appliquer les méthodes suivantes lors de la conception de représentations visuelles :

Processus de conception²⁶⁴ :

- i. Définir clairement les **objectifs** de la représentation (destinataires, attentes, nature des décisions à prendre).
Identifier les **entités et relations pertinentes** sur lesquels l'analyse et le raisonnement vont s'appuyer.
 - ii. Avoir conscience des **difficultés** liées à la nature des informations à représenter (incertitudes et incomplétude des informations, distinction entre informations et hypothèses, mettre en évidence les éléments clefs).
 - iii. Connaître les **biais potentiels** (niveaux de généralité, limites du langage graphique).
 - iv. Mettre en œuvre une **méthode de vérification de la compréhension** du schéma par son destinataire.
-

De même, il résume ainsi les conventions que l'analyste se veut appliquer pour assurer une compréhension des schémas par les pairs et les destinataires :

Recommandations méthodologiques générales²⁶⁵ :

Entités

Les entités importantes sont représentées par des nœuds distincts. Une même entité ne doit apparaître qu'une seule fois par schéma.

Liens

Quatre types de liens doivent être distingués en appliquant les conventions de représentation : confirmés (traits pleins), incertains (traitillés), hypothétiques (pointillés), et absents (couleur spécifique si l'analyste souhaite mettre en valeur un lien non existant, rien sinon)

Les flèches sont utilisées pour décrire des liens directionnels, les couleurs pour décrire des natures de liens (relations familiales, amicales, criminelles, types de transactions ou de communications...)

Cadres

Les cadres sont utilisés pour représenter des groupements d'entités, tels que des cliques, des organisations, des familles... Ils peuvent également décrire des relations, telles que des liens spatiaux (ville, pays), temporels (jour, mois), ou à une même entité (utilisation par plusieurs acteurs d'une même ligne téléphonique, d'un même véhicule).

La couleur du cadre décrira la nature de la relation.

Les cadres pouvant complexifier la lecture, ils ne doivent être utilisés que lorsque les groupements sont pertinents pour l'analyse.

Structure du schéma

Les croisements de traits doivent être évités au maximum, et l'orthogonalité privilégiée pour faciliter l'accès aux informations pertinentes.

²⁶³ Quentin Rossy est Professeur à l'École des Sciences criminelles de Lausanne, responsable de la maîtrise en traçologie et analyse de la criminalité créée conjointement avec l'École de Criminologie de l'Université de Montréal, et titulaire d'un doctorat intitulé « *Méthodes de visualisation en analyse criminelle : approche générale de conception des schémas relationnels et développement d'un catalogue de patterns.* »

²⁶⁴ Quentin Rossy, « La visualisation relationnelle au service de l'enquête criminelle », dans *Les réseaux criminels*, Les Presses de l'université de Montréal., Montréal, 2016, pp. 17-39., p. 25.

²⁶⁵ *Ibid.*, pp. 22-23.

La position des entités dénote cognitivement des relations de proximité, l'analyste doit donc être vigilant quant à la disposition pour éviter de créer des biais (position centrale qui donne une impression d'importance dans le réseau, regroupements positionnels perçus comme des groupes effectifs...)

Documentation du schéma

Les sources doivent être indiquées.

L'ensemble des informations pertinentes doit être représenté, et ce, de manière précise et correcte.

Les conventions doivent être explicitées en légende, et le schéma doit être accompagné d'éléments explicatifs nécessaires à sa juste interprétation.

Enfin, il propose des recommandations méthodologiques spécifiques pour plusieurs grands types de schémas. Deux sont particulièrement intéressants : le schéma de réseau et le schéma de trafic.

Le schéma de réseau : recommandations méthodologiques spécifiques²⁶⁶

- i. Identifier les acteurs du réseau (*i.e.* les acteurs pertinents pour répondre à la problématique). Les personnes constituent le type d'entité principal de l'analyse.
 - ii. Identifier les types de relations pertinentes pour la reconstruction du réseau :
 - Les activités criminelles et autres événements sont représentés par des entités dont le type est défini par l'icône.
 - Les relations interpersonnelles (familiales, amicales, criminelles...) sont représentées par des liens dont la nature est décrite par la couleur. L'épaisseur du trait peut être utilisée pour représenter le nombre de relations.
 - Les relations spatio-temporelles entre les activités sont représentées par des cadres, ce qui permet d'exclure ce type particulier de relations des calculs de centralité établis dans un second temps.
 - Des accolades distinguent les groupes d'entités (séries d'événements, groupes de personnes ou d'objets...)
 - iii. Intégrer les mesures sociométriques de centralité (les plus courantes étant degré, intermédiarité et proximité) :
 - L'une des mesures peut être représentée par un cercle proportionnel à la valeur.
 - Les indicateurs sont inscrits sous la forme d'attributs.
 - iv. Ajouter des entités complémentaires pour représenter les autres objets pertinents pour l'enquête/le renseignement. L'analyse des relations interpersonnelles étant l'objectif principal du schéma de réseau, il convient de restreindre l'ajout d'autres entités à celles qui sont essentielles.
 - v. Les conventions et recommandations générales doivent être suivies.
-

Le schéma de trafic : recommandations méthodologiques spécifiques²⁶⁷

- i. Identifier les acteurs du réseau (*i.e.* les acteurs pertinents pour répondre à la problématique). Les personnes constituent le type d'entité principal de l'analyse.

²⁶⁶ *Ibid.*, p. 29.

²⁶⁷ *Ibid.*, p. 32.

- ii. Représenter les transactions par des liens dirigés entre le vendeur l'acheteur :
 - La couleur du lien permet de distinguer les types de produits ou les voies d'approvisionnement.
 - L'épaisseur du trait est proportionnelle au nombre de produits vendus, aux prix ou à la fréquence des transactions.
 - Le flux du trafic est représenté par une structuration horizontale de gauche à droite.²⁶⁸
 - iii. Exploiter les attributs des liens pour représenter un comparer les déclarations :
 - Les sources (acteurs, témoins, victimes...) sont indiquées en regard de chaque information pour faciliter les comparaisons.
 - Le type de police permet de distinguer le type d'attribut (quantité, prix, source...).
 - iv. Les recommandations du schéma de réseau s'appliquent. En particulier, il est possible d'intégrer les mesures sociométriques de centralité.
 - v. Les conventions et recommandations générales doivent être suivies.
-

b) Réseau comme structure et réseau comme acteur

Les réseaux sociaux se définissent autant par leur structure que par la substance des relations. Si l'analyse formelle des réseaux s'intéresse à la première dimension (étude des structures et effets de structures), des études ethnographiques et cognitives ont abordé la seconde, en documentant l'adoption par les groupes criminalisés de normes communes et d'une identité partagée qui les distinguent des autres réseaux criminels²⁶⁹. À ce jour, les recherches s'intéressent à l'une ou l'autre de ces composantes, mais ne combinent pas ces deux approches. Pourtant, tous les réseaux ne constituent pas des groupes organisés et intentionnels, tandis qu'à l'inverse, toute forme d'organisation sociale constitue *ad minima* un réseau social. Dès lors, comment distinguer les réseaux criminels en tant que structures, *i.e.* qui fonctionnent comme un ensemble d'individus s'associant pour maximiser leurs profits individuels, des réseaux criminels comme acteurs, dans lesquels les protagonistes partagent bien plus qu'une recherche de profit personnel : une appartenance au réseau en tant qu'entité à part entière ?

Tous les réseaux sont le produit de structures de relations sociales. Or, certains résultent aussi de normes et d'identités partagées, concrétisées par des pratiques, des produits et des stratégies permettant au réseau d'exister en tant qu'acteurs à part entière du monde social. Ces pratiques et stratégies, lorsqu'elles sont mises en œuvre par les individus qui composent les

²⁶⁸ L'axe vertical est à éviter, car il induit une perception de hiérarchie souvent inadaptée pour représenter un groupe criminel.

²⁶⁹ Amélie Forget, « Du réseau comme structure au réseau comme acteur : comment les réseaux sociaux se constituent en acteurs collectifs », dans *Les réseaux criminels*, Presses universitaires de Montréal., Montréal, 2016, pp. 83-101.

réseaux, définissent petit à petit le réseau soit en acteur plus centralisé, soit en structure minimale et décentralisée.

Le réseau comme structure est un réseau qui n'existe que par sa structure formelle : un réseau social relie des représentants qui partagent un milieu pratique ou un certain type d'activité, mais dont les choix ne sont pas coordonnés²⁷⁰. Les individus membres de ces réseaux ne sont pas caractérisés par des intérêts, des stratégies ou des objectifs communs, mais par le fait d'être liés *de facto* par les mêmes activités sociales au sein d'un même champ. Le réseau comme structure induit des liens faibles entre les acteurs : il n'est pas le résultat d'actions coordonnées²⁷¹.

Le réseau comme acteur est une forme d'action coordonnée et collective. C'est – comme son nom l'indique – un acteur en soi : les acteurs individuels sont liés par la possibilité d'agir collectivement, des objectifs communs, une action volontaire et stratégique. Le réseau comme acteur entretient donc des liens forts.

Kathryn Sikkink propose de concevoir le réseau comme structure et le réseau comme acteur comme deux idéaux types situés aux pôles d'un continuum²⁷² (figure 15).

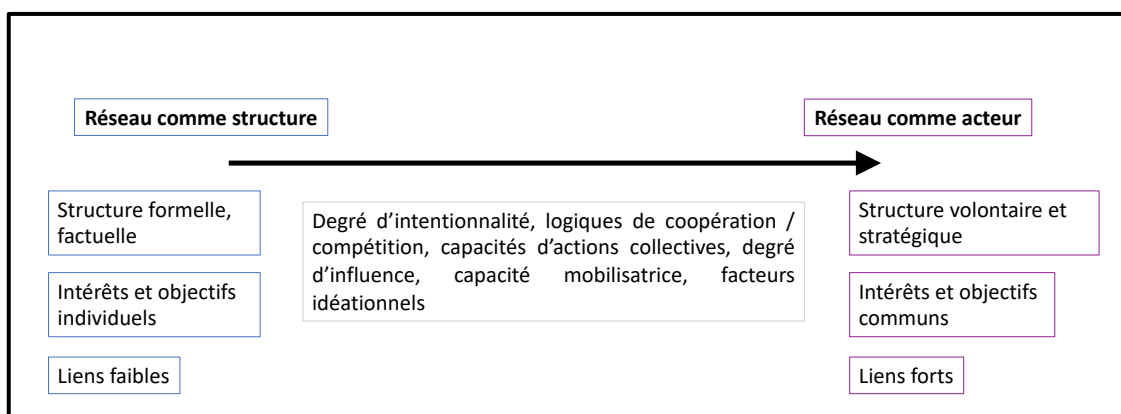


Figure 15 Continuum du réseau comme structure au réseau comme acteur

L'idée du continuum permet d'apprécier le niveau d'intentionnalité d'un réseau comme acteur de façon relative par rapport aux autres, et de hiérarchiser différentes cliques au sein d'un même réseau.

²⁷⁰ *Ibid.*

²⁷¹ *Ibid.*

²⁷² Kathryn Sikkink, « The power of networks in international politics », dans *Networked politics: agency, power, and governance*, Ithaca, Cornell University Press, 2015, pp. 228-247.

Pour qu'un réseau comme acteur existe, il doit disposer d'une structure de relations, d'un enjeu suffisamment important pour lier les individus et de « règles du jeu », c'est-à-dire de normes, formelles ou informelles qui font sens pour les agents.

Quel que soit l'endroit où se place un réseau donné sur ce continuum, les ressources, ou capitaux, sont valorisés au sein du réseau : leur valeur est reconnue par les autres acteurs et permet à leur détenteur d'exercer une influence, un pouvoir, et donc d'exister dans le champ que détermine le réseau²⁷³.

II.4 Éléments de méthodes analytiques appliquées pour la modélisation de réseaux

Les éléments évoqués dans les sections précédentes de ce chapitre se rapportent à l'étude d'individus et de groupes sociaux. Les méthodes d'analyse de réseaux qualitatives et quantitatives s'appliquent ainsi sur des réseaux sociaux, qui peuvent être modélisés sous forme de sociogrammes c'est-à-dire de graphes représentant uniquement des individus (les sommets du graphe) et les relations les unissant (les arêtes du graphe).

Lorsque la donnée brute le permet, c'est-à-dire lorsque l'analyste (ou la chercheuse) dispose directement d'informations sur l'existence ou non de liens entre les différents acteurs étudiés, la construction du sociogramme ne pose pas de difficultés. En fonction de la structure, du formatage et de la quantité de données initiales et selon les besoins et les outils à disposition, le graphe relationnel peut être obtenu soit manuellement, soit de manière automatisée (ou semi-automatisée). Dans le premier cas, l'analyste crée un à un les sommets représentant les acteurs et les liens qui les unissent à l'aide d'outils²⁷⁴. Dans le second cas, l'analyste doit créer un fichier de données sous forme « d'*edge list* »²⁷⁵ importé dans un logiciel²⁷⁶ qui construira le graphe correspondant.

Cependant, il n'est pas rare que les données brutes ne soient pas uniquement constituées de liens entre les différents acteurs : ces derniers peuvent être liés directement, ou indirectement par le biais d'entités et de liens intermédiaires, tels que le fait que deux personnes travaillent

²⁷³ Amélie Forget, « Du réseau comme structure au réseau comme acteur : comment les réseaux sociaux se constituent en acteurs collectifs », *op. cit.*

²⁷⁴ Tels que le logiciel *Stemic*, développé par la société Intactile Design qui finance cette thèse.

²⁷⁵ C'est-à-dire un tableau contenant sur chaque ligne une entité source, un lien et une entité destination (ainsi que les attributs de ces trois objets s'ils en ont).

²⁷⁶ Tel que le logiciel *Tulip*, développé par le LaBRI qui a collaboré au développement de la méthode d'extraction du réseau social présentée au chapitre IV.

dans la même école, ou assistent aux mêmes événements. Les graphes relationnels représentant ce type de données sont dits « multivariés » car ils contiennent plusieurs types d'entités (contrairement aux sociogrammes).

Dans ce dernier cas, il est nécessaire pour appliquer les méthodes d'ARS d'extraire le sociogramme du réseau social depuis le graphe multivarié. Cette opération est dite de « transformation de modèle ».

a) Transformations de modèles : application de la théorie des graphes

Les méthodes classiques de transformation de modèles permettent l'extraction d'un réseau social N , *i.e.* d'un graphe *one-mode*²⁷⁷, depuis un graphe multivarié G . Ces méthodes s'appliquent usuellement lorsque G est un graphe *two-mode* (aussi appelé « graphe biparti »). Il comporte alors deux types d'entités distinctes. Les arêtes représentent des liens inter-entités uniquement, c'est-à-dire qu'il n'existe pas de liens reliant directement deux entités de même type²⁷⁸. La figure 16 ci-dessous montre un exemple de graphe biparti.

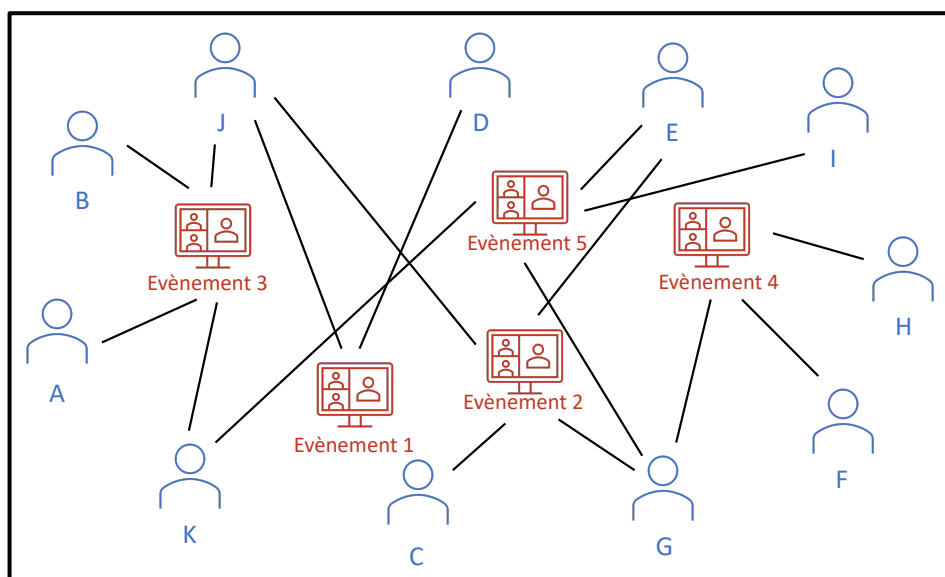


Figure 16 Exemple de graphe *two-mode* G représentant 11 individus et les événements auxquels ils ont assisté les 6 derniers mois.

Dans cette situation, les individus ne sont jamais liés entre eux directement, ils sont uniquement connectés aux événements auxquels ils ont participé. Le réseau social N obtenu depuis G lie les personnes ayant co-participé à au moins un événement (figure 17). Dans ce cas souvent rencontré, on dit parfois que N est obtenu par projection du graphe G . Les liens du

²⁷⁷ C'est-à-dire ne contenant qu'un seul type d'entités – des individus (ou groupes d'individus) pour un sociogramme.

²⁷⁸ Stephen P. Borgatti, Martin G. Everett et Jeffrey C. Johnson, *Analyzing social networks*, *op. cit.*

réseau social peuvent être pondérés pour, notamment dans cet exemple, comptabiliser le nombre d'événements auxquels les individus ont participé, ce qui permet de conserver une partie de l'information originale.

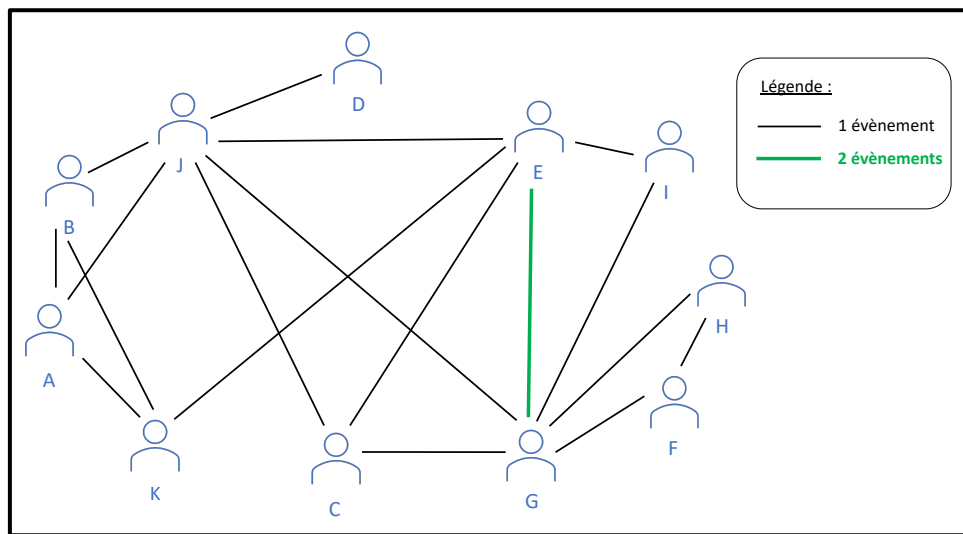


Figure 17 Réseau social N extrait du graphe G de la figure 16.

Ces transformations, ou projections, sont décrites avec le formalisme suivant : soit un graphe biparti $G = (V, E)$ non orienté et sans boucle avec $V = \{v_0, \dots, v_n\}$ l'ensemble des sommets et $E = \{e_0, \dots, e_m\}$ l'ensemble des arêtes. La transformation de modèle consiste dans ce cas à induire à partir de G un réseau social N , c'est-à-dire à calculer à partir de $G = (V, E)$ un graphe $N = (P, F)$ où les sommets $P \subset V$ sont des individus et où les nouvelles arêtes $F \subset E$ lient directement les individus connectés aux mêmes sommets de l'autre type dans G ²⁷⁹.

b) Traitement automatisé du langage naturel : principe général de grammaire formelle

Le Traitement automatisé du Langage naturel (TALN) est une discipline à la frontière entre la linguistique et l'intelligence artificielle, rassemblant l'ensemble des méthodes et outils permettant à un programme informatique de comprendre le langage tel qu'il est construit.

Le terme « grammaire » désigne une théorie des structures linguistiques dont la pièce maîtresse est la syntaxe – distincte de la sémantique. Une grammaire est ainsi un formalisme permettant de définir une syntaxe et donc un langage formel, tel qu'un ensemble de mots

²⁷⁹ Bruno Pinaud, Maud Benichou et Guy Melançon, « Extraction d'un réseau social criminel par transformation d'un graphe d'enquête multivarié », Conférence Extraction et Gestion des Connaissances (EGC), Lyon, 2023. Article accepté pour publication en format long par le comité de lecture.

admissibles sur un alphabet donné. Elle rend compte de toutes et seulement les phrases grammaticales – *i.e.* syntaxiquement bien formées – qui lui sont associées. Pour ce faire, une grammaire est conçue sous la forme d'un mécanisme génératif, c'est-à-dire qu'en partant d'un axiome de départ et d'un ensemble de règles de réécriture, elle doit engendrer toutes les phrases syntaxiquement correctes possibles et imaginables.

Le vocabulaire en grammaire formelle peut être décomposé et défini comme suit :

1. Un ensemble fini de symboles terminaux : les plus petits éléments du langage formalisé et donc irréductibles.
2. Un ensemble fini de symboles non terminaux, résultant de la combinaison de terminaux et non terminaux selon les règles de réécriture qui régissent le langage.
3. Un ensemble de règles de réécriture régissant les modalités d'association des terminaux et non terminaux.

L'intégration de ces principes à la méthodologie d'analyse est décrite dans le chapitre V « méthodologie générale : du graphe multivarié au réseau social ».

CHAPITRE III CAS D'USAGE ET DONNÉES D'ÉTUDE : PRÉSENTATION D'UNE ENQUÊTE D'ENVERGURE ET STRUCTURATION DES DONNÉES

« Dans la conception de l'analyste criminel, les données recueillies constituent le matériel brut. »

Olivier Ribaux²⁸⁰.

Ce travail de recherche se voulant exploratoire en termes de méthodologie, il est concrétisé par l'application des méthodes proposées à un cas d'usage réel, fourni par le Service central de Renseignement criminel de la Gendarmerie nationale (SCRC-GN), que nous nommerons le « dossier K ». Il s'agit d'une enquête conjointe des Divisions de la Criminalité organisée (DCO) des Sections de Recherche (SR) de Paris XXème et de Marseille. Les investigations s'étendent de 2015 à 2018, mettant à disposition un volume de données considérable. Ces investigations ont permis de démanteler une importante organisation de trafic de fonds sur le territoire français et de blanchiment à l'étranger.

III.1 Accès aux données : spécificités administratives et pratiques

a) Aspects administratifs

Cette affaire, dont le jugement a été rendu en octobre 2019, a depuis cette date « autorité de la chose jugée²⁸¹ », rendant ainsi possible son étude pour ces travaux de recherche. Cependant, de par son caractère réel et donc sensible, des règles de confidentialité doivent être respectées pour bénéficier de l'accès au dossier. Aucune donnée nominative ne peut être sortie par la doctorante des locaux de la GN ; ces dernières ne peuvent être consultées, traitées, analysées et exploitées que dans l'enceinte de la GN, en conséquence de quoi toute diffusion

²⁸⁰ Olivier Ribaux, *Police scientifique, op. cit.*, p. 369.

²⁸¹ L'autorité de la chose jugée peut être définie comme une force exceptionnelle conférée par la loi aux décisions juridictionnelles, qui une fois prononcées bénéficient du principe de l'immutabilité interdisant de remettre en cause ce qui a été définitivement jugé.

ou publication devra être précédée d'une anonymisation systématique de l'ensemble du jeu de données.

Pour ces mêmes raisons, les processus d'autorisations d'accès sans censure au dossier et aux données qu'il contient ont duré plusieurs mois après le début des travaux, conformément au processus prévu par la circulaire n°30053 du 3 septembre 2019 relative à l'export de données à destination de la recherche scientifique²⁸². Avec l'appui du Lieutenant-Colonel Barlatier, chef de la Division du Renseignement (DR) du SCRC, l'habilitation a ainsi été accordée le 13 mars 2020, date à laquelle aurait dû débiter l'étude du dossier.

b) Aspects pratiques

Cependant, suite au confinement dû à la covid-19, la Gendarmerie n'a pu accueillir de personnel extérieur jusque fin mai 2020, retardant ainsi l'accès aux données²⁸³. Initialement, ces dernières étaient consultées dans les locaux de la Section de Recherche du XXème arrondissement de Paris. En effet, les enquêteurs qui avaient réalisé les investigations ont été d'une aide précieuse dans la découverte et la compréhension des faits et de l'enquête. Après huit mois d'exploration des données, la décision a été prise de transférer l'ensemble du dossier au SCRC, permettant ainsi de s'appuyer sur les connaissances et l'expérience des analystes en renseignement criminel qui y travaillent.

Ainsi, en juin 2020, à la demande du SCRC, l'ensemble des pièces de procédure en version papier originale nous a été rendu accessible par la SR de Paris XX^{ème}.

La première phase d'exploration a été compliquée, par manque d'expérience et de méthode face à des données non numérisées et non structurées. Durant les trois premiers mois, une méthode d'acquisition pertinente et efficiente a été recherchée. En septembre 2020, la décision a été prise de consulter l'un des analystes de la DCO²⁸⁴. À la demande de la doctorante, la matrice utilisée par les analystes de la SR pour la saisie des données issues de procès-verbaux a été mise à disposition. Dans un premier temps, la matrice vierge de toute donnée a été obtenue : l'objectif était de travailler avec une structure cohérente nous permettant de saisir les

²⁸² Cf. annexe 4 : « Circulaire ministérielle n°30053 du 3 septembre 2019 relative à l'export de données à destination de la recherche scientifique ».

²⁸³ Pour ces raisons, l'Association nationale de la Recherche et des Technologies (ANRT) a accordé à de nombreux doctorants en thèse Cifre un délai supplémentaire, permettant de continuer les travaux au-delà des 36 mois initialement prévus. Dans notre cas, une prolongation du contrat Cifre de trois mois a été admise.

²⁸⁴ Il n'a malheureusement pas été possible d'échanger avec l'analyste ayant traité le dossier. La discussion a eu lieu avec l'un de ses collègues.

informations contenues dans les pièces de procédure originales de manière efficace et pertinente.

Par la suite, compte tenu du volume de données à saisir manuellement, des aspects administratifs susmentionnés ayant retardé l'accès au dossier et de la date limite fixe, contractuellement parlant, de rendu du manuscrit de thèse²⁸⁵, une nouvelle stratégie a été adoptée en décembre 2020. Une procédure de demande d'autorisation d'exploitation du fichier déjà saisi par l'analyste qui avait traité l'affaire a été entamée. Afin d'éviter tout biais, seuls les fichiers de saisie originaux ont été demandés, c'est-à-dire vierges de tout traitement ou analyse. La procédure a abouti fin février 2021, date à laquelle la matrice d'enquête a été obtenue.

III.2 Description des données à disposition : forme, structure et contenu

a) Résumé de l'enquête²⁸⁶

D'après le réquisitoire définitif du xx mars 2019 et l'arrêté du xx avril 2019²⁸⁷ : L'enquête a révélé un vaste réseau international de transport et blanchiment de fonds dirigé depuis le Maroc, *via* le système de l'« *hawala* », système traditionnel informel de paiement datant du Moyen-Âge initialement utilisé pour ne pas voyager avec de l'argent et donc éviter de se faire détrousser par les bandits de grands chemins.

Basé sur la confiance et imposant un faible taux de change (ou plus exactement taux de transfert), ce modèle financier fonctionne par compensation sans aucun transfert d'argent liquide : pour faire circuler de l'argent du Maroc vers la France par exemple, le client se présente au « *saraf* » au Maroc et lui remet en liquide la somme d'argent qu'il souhaite transférer en France. Il reçoit en échange un « *token* », morceau de papier sur lequel est inscrit un code. Le *saraf*, mi-courtier mi-banquier informel, contacte alors son correspondant, un autre *saraf* en France, et lui indique le montant à remettre ainsi que le code dont dispose le client. Lorsque ce dernier arrive en France, il se présente au *saraf* français, et récupère la somme d'argent liquide en euros²⁸⁸. Ses applications vont aujourd'hui bien au-delà du transfert de fonds pour des particuliers : de par l'impossibilité de tracer l'argent comme la transaction, ses

²⁸⁵ Après prolongation, l'échéance est fixée au 27 novembre 2022, soit deux mois après la fin du contrat Cifre.

²⁸⁶ Dans une démarche exploratoire, scientifique et rigoureuse, le présent travail de recherche a été effectué « à l'aveugle » de bout en bout. Le résumé de l'affaire a donc été rédigé en date du 14 novembre 2022, après avoir pris connaissances du réquisitoire définitif du x mars 2019 et de l'arrêté du x avril 2019 le 11 novembre 2022, lorsque les derniers résultats des analyses ont été extraits.

²⁸⁷ Les dates exactes ne sont volontairement pas données, conformément à la circulaire ministérielle n°30053 du 3 septembre 2019.

²⁸⁸ Luc Leroux, « Un tentaculaire réseau de blanchiment d'argent en procès à Marseille », *Le Monde*, 29/05/2019

utilisations dans les milieux criminels sont nombreuses²⁸⁹ : passeurs clandestins, paiement des fournisseurs de marchandises aux trafiquants, blanchiment d'argent sale...

Les utilisateurs de ce trafic de fonds et de ce réseau de blanchiment étaient des clients particuliers ainsi que des réseaux de trafics de stupéfiants français, espagnols et belges. Les trafiquants s'adressaient à un *saraf* près de leur domicile, à qui ils déposaient l'argent en liquide et qui était ensuite chargé de transférer l'argent aux exportateurs de stupéfiants au Maroc en utilisant le système de l'*hawala*. Les surveillances et écoutes téléphoniques ont permis d'identifier les membres de la famille M.²⁹⁰, avec à leur tête A.M., comme étant les *saraf* marocains qui pilotaient l'activité des collecteurs de fonds implantés en France et en Belgique, mobiles jusqu'en Espagne.

Ces *saraf* marocains étaient en lien avec une famille de *saraf* belge, la famille B., avec à sa tête le patriarche N.B., largement secondé par son fils, T.B. Les M. émettaient, par téléphone ou messages cryptés, les ordres de collecte et de versement.

A.M. était directement en contact avec les clients du réseau, désignés par plusieurs codes, tels que « code 44 », « code T », « code R » qui semblaient correspondre à plusieurs réseaux de trafiquants en France, en Espagne et en Belgique.

Les B. disposaient de plusieurs collecteurs pour répondre aux ordres de collecte et de versement des M., notamment D.B. (second fils de N.B.), I.C.D. (gendre de N.B.) et J.S. Ils disposaient d'un lieu de stockage et de transit des fonds situé dans la ville d'Anderlecht en Belgique. En France, J.S. obéissait aux ordres de collecte donnés par N. et T.B. Si certains contacts de J.S. s'avéraient être de simples collecteurs, plusieurs autres semblaient être des trafiquants de stupéfiants, ou du moins des collecteurs de réseaux de trafiquants. Les filatures de J.S. ont établi qu'il déposait souvent l'argent collecté dans une épicerie du 17^{ème} arrondissement de Paris. De très nombreuses allées et venues dans ce commerce ont été constatées, ayant tant pour objet le dépôt d'importantes sommes d'argent liquide que le retrait de sommes plus faibles par des particuliers. Il s'agissait là du *saraf* français, qui tenait une véritable banque clandestine en plein Paris.

Les fonds étaient ensuite transférés vers plusieurs individus, dont un investisseur paraissant jouer un rôle prégnant, et vers les Pays-Bas. Des flux financiers vers les Émirats Arabes Unis (EAU) étaient également mis en évidence, notamment via un trader implanté à Dubaï, lui-même en contact avec A.M. L'existence de plusieurs autres traders aux EAU œuvrant pour A.M. était également mise en évidence. Les opérations effectuées concernaient des

²⁸⁹ Éric Vernier, *Techniques de blanchiment et moyens de lutte*, Paris, Dunod, 2005, 304 p.

²⁹⁰ Initiales fictives conformément à la circulaire ministérielle n°30053 du 3 septembre 2019.

sommes de plusieurs centaines de milliers d'euros. On peut lire, « *Dans le cadre de ce réseau criminel structuré, le montant des espèces collectées depuis le début de l'enquête excédait 70 000 000 €, cette somme représentant uniquement les remises formellement identifiées et le chiffre réel étant sans doute bien supérieur à ce montant.* » Enfin, il est fait mention du fait que certains acteurs du dossier ont été interpellés en possession de grandes quantités de stupéfiants (au total, 785 kg de cannabis et 80 kg de cocaïne ont été saisis), confirmant la nature frauduleuse des fonds alimentant le circuit, à savoir les recettes de trafics de stupéfiants.

b) Procès-verbaux (PV) de procédure : la donnée non structurée

Cette section décrit la composition du « dossier K » qui constitue le cas d'usage. L'ensemble du dossier dans sa version papier originale nous a été rendu accessible. Ce dernier est divisé en trois parties, nommées respectivement comme suit : « K**** IDF – Envois I à V », « K**** IDF CRT – Envois I à V », « K**** IDF – OP LOLA A et B »²⁹¹.

i. « K**** IDF – Envois I à V »

Cette partie du dossier contient des documents de tous types relatifs à l'enquête.

La première pièce du dossier est l'arrêt n° xxx/2019 du x avril 2019 de la Chambre de l'instruction de la Cour d'Appel d'Aix-en-Provence ayant siégé en Chambre du Conseil pour les débats le x mars 2019 pour le prononcé de l'arrêt le x avril 2019²⁹².

Ce document reporte, « vue l'information suivie au tribunal de grande instance de Jirs-Marseille contre [34 individus] des chefs de : blanchiment d'importation de stupéfiants en bande organisée, blanchiment du produit de trafic de stupéfiants, participation à une association de malfaiteurs en vue de la préparation de crimes ou de délits punis de 10 ans d'emprisonnement », le déroulement des débats, l'ensemble des attendus, les motifs de décision, la décision en la forme et au fond.

Viennent ensuite un grand nombre de sous-dossiers composés des éléments suivants : le bordereau d'envoi de la commission rogatoire mentionnant la désignation des pièces, la commission rogatoire²⁹³ (CR), les procès-verbaux (PV) décrivant les actes d'enquêtes effectués

²⁹¹ Données nominatives et numéros de référence des pièces volontairement non reportés en vertu de la circulaire ministérielle n°30053 du 3 septembre 2019.

²⁹² *Idem* pour les dates.

²⁹³ La commission rogatoire est l'acte juridique par lequel un juge d'instruction délègue ses pouvoirs à un autre magistrat ou à un officier de police judiciaire (OPJ), afin de faire procéder à sa place à un ou plusieurs actes d'information. Elle est prévue par l'article 151 du Code de Procédure Pénale (CPP).

(investigation, mise sous surveillance, réquisition, retranscription d'interception, mise sous scellé, saisine, synthèse), les scellés s'ils existent et si, le cas échéant, leur format permet de les stocker avec la procédure, ainsi que l'inventaire des pièces à conviction.

Ces documents contiennent des données nominatives et personnelles, telles que des identités, dates de naissance, filiations, photos d'individus ou d'objets mentionnés dans la procédure, copies de documents, plaques d'immatriculation de véhicules, adresses et numéros de téléphone, des faits datés, des lieux, des montants d'argent...

Les sous-dossiers peuvent concerner une ou des personne(s) précises, un ou des véhicule(s) en particulier, un ou des lieu(x), une période temporelle spécifique, un ou des numéro(s) de téléphone mis sous écoute, etc., selon les éléments et besoins de l'enquête en cours. Ainsi, les pièces sont triées par sous-dossier en fonction des objets auxquelles elles font référence plutôt que chronologiquement.

Enfin, la dernière pièce du dossier est le « Réquisitoire définitif de disjonction, de requalification, de non-lieu partiel et de renvoi devant le tribunal correctionnel », rédigé au Parquet de Marseille le x mars 2019, à l'intention du Juge d'instruction par deux Vice-Procureurs pour le Procureur de la République. Il s'agit d'une pièce de procédure écrite par laquelle le Procureur de la République donne au Juge d'instruction, à la fin de l'instruction, sa position sur les suites à donner à l'affaire, en vertu de l'article 175 du Code de procédure pénale (CPP). Ce document, constitué de 257 pages, expose, « vues les pièces de l'information suivie contre [36 individus]²⁹⁴, les chefs d'accusation de :

- Blanchiment d'importation de stupéfiants en bande organisée,
- Blanchiment du produit de trafic de stupéfiants,
- Recel d'importation de stupéfiants en bande organisée,
- Recel du produit de trafic de stupéfiants,
- Complicité de blanchiments aggravés,
- Association de malfaiteurs,
- Blanchiment en bande organisée,
- Non justification de ressources,
- Recel de vol.

Il se compose d'un sommaire, d'un rappel des origines de l'affaire, d'un détail des acteurs du réseau, des éléments à charge et à décharge, d'éléments de personnalité et de

²⁹⁴ L'arrêt n° xxx/2019 du x avril 2019 mentionne bien 34 individus, contre 36 pour le réquisitoire définitif de disjonction, de requalification, de non-lieu partiel et de renvoi devant le tribunal correctionnel.

patrimoine des individus concernés, d'un détail des scellés et saisies pénales, des réquisitions finales, ainsi que d'un tableau récapitulatif en annexe.

ii. « K**** IDF CRT – Envois I à V »

Cette partie du dossier contient l'intégralité des documents et pièces de procédure relatifs aux commissions rogatoires techniques (CRT). Les CRT sont des CR spécifiques qui concernent les actes d'information dits « techniques », tels que la mise sur écoute (ou interception) de lignes téléphoniques, l'installation de balises de localisation sur des véhicules, la sonorisation de lieux ou de véhicules, etc.

iii. « K**** IDF – OP LOLA A et B »

Cette troisième et dernière partie du dossier se concentre sur les investigations menées autour des activités en rapport avec l'affaire d'une personne en particulier que nous nommerons Lola Ruffaut²⁹⁵. Les documents contenus dans cet élément du dossier sont similaires à ceux décrits plus hauts. En préambule du PV de synthèse, il est indiqué que « les investigations menées dans le cadre de la procédure initiale (PV xxx/xxxx/2015) ont permis de déterminer que la dénommée [Lola Ruffaut] exerçait une activité illicite pour le compte de son père, [Christian Ruffaut]²⁹⁶ demeurant au Maroc ».

Sont mentionnés, tels que révélés par l'enquête, les personnes incluses dans ces activités ainsi que leurs degrés d'implication.

c) Matrices d'analyse : la donnée structurée

1. Description de la structure matricielle

Dans cette section est décrite la matrice de données utilisée lors de l'enquête telle que renseignée par l'analyste criminel, avant toute analyse à proprement parler. Il s'agit dans notre cas des données brutes. Cette matrice est la structure utilisée par l'ensemble de la communauté d'analystes en gendarmerie nationale, et permet de saisir des données relationnelles sous forme d'*edge list*. Cela signifie que chaque ligne du fichier de données correspond à un lien, révélé

²⁹⁵ Pseudonyme attribué dans le respect de l'anonymisation en vertu de la circulaire ministérielle n°30053 du 3 septembre 2019.

²⁹⁶ Pseudonyme attribué dans le respect de l'anonymisation en vertu de la circulaire ministérielle n°30053 du 3 septembre 2019.

lors de l'enquête, entre deux entités. C'est pourquoi nous abordons dans l'ensemble du travail ces données par la théorie des graphes, c'est-à-dire en travaillant avec un ensemble d'entités – les nœuds du graphe – et de liens les unissant – les arêtes du graphe.

La matrice d'enquête est composée de plusieurs onglets, dont le premier, « documents » permet de recenser les différentes pièces de procédure (référence du PV, numéro de pièce, unité d'enquête, type de document, enquêteur, traitement). Les onglets suivants, au nombre de neuf, correspondent aux types d'entités depuis lesquels les relations ont été observées lors de l'enquête : « personnes », « personnes morales », « télécom », « lieux », « objets », « véhicules », « comptes », « armes », « événements ». Pour cette enquête, seuls les six premiers onglets ont été utilisés par l'analyste. L'ensemble des données (*i.e.* l'ensemble des liens) est réparti sur ces onglets en fonction du type de l'entité depuis laquelle le lien est observé, ce qui permet, pour chacun d'entre eux, de renseigner les attributs pertinents de l'entité source (par exemple date de naissance pour des personnes, immatriculation pour des véhicules, etc...). La structure générale est commune à ces onglets. La donnée relationnelle contient toujours les informations suivantes : entité source, icône descriptive de l'entité source²⁹⁷, attributs de l'entité source (de 4 à 7 colonnes selon le type d'entité), libellé du lien, type de lien, attributs du lien (solidité, largeur, date, heure), entité destination, icône associée au type d'entité destination, références de la pièce de procédure contenant l'information.

Les tableaux 3, 4 et 5 montrent la structure utilisée pour stocker les données, en prenant pour le tableau 3 l'exemple de l'onglet « personnes ». Le tableau 6 indique la liste détaillée des différentes entités existantes dans les données.

Tableau 3 **En-têtes de colonnes concernant l'entité source dans l'onglet « personnes »**

NOM PRENOM	ICONE SOURCE	Date de naissance	Lieu de naissance	Antécédents	Profession	Statut	Divers
---------------	-----------------	----------------------	----------------------	-------------	------------	--------	--------

Tableau 4 **En-têtes de colonnes concernant le lien, communes à tous les onglets**

LIBELLE LIEN	TYPE LIEN	SOLIDITE LIEN	LARGEUR LIEN	DATE LIEN	HEURE LIEN	REF PV	UPVA
--------------	-----------	---------------	--------------	-----------	------------	--------	------

Tableau 5 **En-têtes de colonnes concernant l'entité destination, communes à tous les onglets**

DESTINATION	ICONE DESTINATION
-------------	----------------------

²⁹⁷ L'icône associée à chaque entité représente la nature de l'entité (homme, femme, voiture, camion, téléphone...) et permet de faciliter la visualisation relationnelle.

Tableau 6 **Liste initiale des différentes entités (information stockée dans la colonne « icônes »)**

Compte	Carte de crédit bancaire (dos)	Téléphone
Restaurant	Drogue	Téléphone portable
Café	Homme	Courrier électronique
Magasin	Femme	Serveur
Lieu	Anonyme	Voiture
Appartement	Garçon	Moto
Aéroport	Fille	4X4
Maison	Clan	Fourgon
Garage	Bureau	Camion-plateau
Document	Carte SIM	

2. Description des entités et de leurs attributs

Concernant le contenu, le volume total de données à disposition est initialement de 2 396 liens, pour 1 736 entités différentes. Les données nécessitant toujours une importante phase de préparation, ou pré-traitement, nous avons choisi de présenter ici les volumes de données sur lesquels se base l'analyse à proprement parler, c'est-à-dire après application des pré-traitements décrits dans la section IV.1 « Préambules de l'analyse : pré-traitements et traitements des données ». La structure reste cependant inchangée. Le matériel de travail contient ainsi 2 357 liens pour 1 747 entités.

Le tableau 7 indique la répartition des entités par type. 38,75% des entités considérées sont de type « PERSONNE ».

Tableau 7 **Répartition des entités par type**

Type d'entité	Nombre d'entités
BANQUE	1
LIEU	500
OBJET	7
PERSONNE	677
PERSONNE_MORALE	104
TELECOM	195
VEHICULE	263
Total général	1747

Comme évoqué plus haut, pour chaque type d'entité, différents attributs sont considérés dans la matrice originale. Ces derniers sont décrits dans les tableaux 8 à 13 ci-dessous, accompagnés d'un exemple²⁹⁸, et du taux de remplissage de chaque colonne.

Tableau 8 Colonnes concernant l'entité source dans l'onglet « personnes », exemple de données et taux de remplissage

Colonnes	NOM PRENOM	ICONE SOURCE	Date de naissance	Lieu de naissance	Antécédents	Profession	Statut	Divers
Exemple	LEDANT Maurice	Homme	03/08/1974	VENISSIEUX (69)	Stupéfiants, violence moins de 8 jours, conduite sans permis, destruction document ou objet pouvant faciliter la découverte d'un crime ou délit, recel de vol	/	/	Surnom "Darty"
Taux de remplissage	100%	100%	47,99%	41,19%	15,36%	1,22%	0%	18,50%

Tableau 9 Colonnes concernant l'entité source dans l'onglet « personnes morales », exemple de données et taux de remplissage

Colonnes	RCS (ou Enseigne)	ICONE SOURCE	RCS	Type	Activité	Divers
Exemple	SARL GLOTRADE	Bureau	/	/	/	Commerce de gros
Taux de remplissage	100%	100%	0%	0,85%	5,08%	11,86%

Tableau 10 Colonnes concernant l'entité source dans l'onglet « télécom », exemple de données et taux de remplissage

Colonnes	NMR TPH	ICONE SOURCE	Opérateur	Type	Surveillance	IMEI	Statut	Divers
Exemple	+33612345678	Carte SIM	/	/	/	ACE	/	Désactivé
Taux de remplissage	100%	100%	10,67%	0%	0%	20% ²⁹⁹	0%	10,67%

Tableau 11 Colonnes concernant l'entité source dans l'onglet « lieux », exemple de données et taux de remplissage

Colonnes	NOM LIEU	ICONE SOURCE	Ville	Code postal	Type	Valeur	Statut	Divers
Exemple	12 rue Deschamps	Lieu	Paris	75013	/	/	/	2eme étage
Taux de remplissage	100%	100%	80%	80%	0%	0%	0%	13,33%

²⁹⁸ Données anonymisées : pour des raisons de confidentialité, seules des informations fictives sont ici renseignées.

²⁹⁹ Dans le tableau de données, les informations renseignées dans cette colonne ne correspondent pas en réalité au format attendu pour un numéro IMEI. Ces données sont donc inutilisables, dans la mesure où il n'est pas possible de savoir à quoi elles correspondent.

Tableau 12 Colonne concernant l'entité source (dans l'onglet « objets »), exemple de données et taux de remplissage

Colonnes	NATURE	ICONE SOURCE	Type	Numéro	Marque	Modèle	Couleur	Statut	Divers
Exemple	Certificat de conformité	Document	/	/	/	SQ5	/	/	/
Taux de remplissage	100%	100%	18,18%	9,09%	0%	18,18%	0%	0%	18,18%

Tableau 13 Colonne concernant l'entité source (dans l'onglet « véhicules »), exemple de données et taux de remplissage

Colonnes	N° IMMAT	ICONE SOURCE	Marque	Modèle	Type	Couleur	Statut	Divers
Exemple	BN-123-CH	Voiture	Mercedes	Classe B	/	Blanche	/	En lien BM express
Taux de remplissage	100%	100%	96,97%	89,90%	5,05%	30,30%	0%	12,12%

Tableau 14 Taux de remplissage des attributs des entités dans la matrice, par onglet et au total sur l'ensemble des données

Onglet	Taux de remplissage
« Personnes »	20,53%
« Personnes morales »	4,45%
« Télécom »	6,89%
« Lieux »	28,89%
« Objets »	10,00%
« Véhicules »	39,22%
Matrice intégrale	20,02%

Le tableau 14 montre que seulement 20,02% des attributs des entités sont renseignés, ce qui est relativement faible. Il ne sera ainsi pas possible d'impliquer les attributs des entités de manière générique dans la modélisation. En effet, le biais résultant du grand nombre de données manquantes serait trop important, et risquerait de dégrader la précision du modèle ainsi obtenu. En revanche, dans certains cas particuliers et *a priori* lors des phases de traitement des données analysées, lorsque l'on s'intéresse à des entités précises, si l'information existe, elle pourra être utilisée de manière ponctuelle. À titre d'exemple, si deux véhicules sont désignés par des plaques d'immatriculation incomplètes et identiques, afin de déterminer s'il s'agit du même véhicule, il sera possible pour ces deux entités d'aller regarder, si elles existent, les informations sur leurs attributs (marque, modèle, couleur) et d'en tenir compte.

3. Relations décrites et informations associées

Le fichier de données étant sous forme *d'edge list*, chaque ligne correspond à une relation entre deux entités, qu'elles soient ou non de même type. Les relations sont initialement réparties dans les six onglets utilisés dans ce cas d'usage par l'analyste en fonction du type de l'entité depuis laquelle le lien est observé.

Le tableau 15 ci-dessous montre comment sont répartis les liens dans les différents onglets. Pour 75,04% des relations jugées pertinentes, le lien est observé depuis une entité de type « personne ».

Tableau 15 Répartition des liens par type d'entité source

Type d'entité source	Volume de liens établis depuis ces entités
LIEU	38
OBJET	6
PERSONNE	1774
PERSONNE_MORALE	166
TELECOM	120
VEHICULE	253
Total general	2357

Les liens, comme les entités, sont caractérisés par leur libellé et leurs attributs, au nombre de cinq :

- Le type de lien (« autre », « utilisateur », « localisation », « téléphonique », « propriété », « familial - amical », « criminel », « bancaire », ou « professionnel »), qui indique la nature de la relation ;
- La solidité du lien (« Confirmé » ou « Non Confirmé »), qui indique le niveau de fiabilité de la relation ;
- La largeur du lien³⁰⁰, qui permet aux gendarmes de définir l'épaisseur souhaitée du lien lorsque les données sont visualisées graphiquement après import dans leur logiciel d'analyse, *IBM Security i2 Analyst's Notebook*³⁰¹ ;
- La date du lien ;
- L'heure du lien ;

En outre, deux colonnes indiquent la pièce de procédure dont vient l'information : la première, « Ref PV » est une référence propre à l'analyste, la seconde, « UPVA » est la référence complète du procès-verbal. Le tableau 16 indique la répartition des liens par type.

³⁰⁰ Dans ce cas d'usage, la valeur renseignée par l'analyste est identique pour toutes les relations.

³⁰¹ Développé et commercialisé par IBM, *IBM Security i2 Analyst's Notebook* est un outil d'analyse visuelle aidant à transformer les données en informations, au sens de Ratcliffe (*cf.* fig. 3). Cette solution est notamment utilisée afin de « mettre en évidence les connexions et les modèles cachés dans les données ». Source : <https://www.ibm.com/fr-fr/products/i2-analysts-notebook>.

Tableau 16 Répartition des liens par type

Type de lien	Nombre de liens
AUTRE	304
BANCAIRE	12
CRIMINEL	96
FAMILIAL - AMICAL	173
LOCALISATION	815
PROFESSIONNEL	116
PROPRIETE	367
TELEPHONIQUE	70
UTILISATEUR	404
Total general	2357

Les liens de localisation, soit 34,58% des relations renseignées, sont les plus fréquents. Viennent ensuite les liens d'utilisation (17,14%) et de propriété (15,57%). Il est important de noter que le type « autre » contient plus de 10%³⁰² des relations. Ce type de lien n'apportant aucune information en soi, un travail sur les libellés devra être effectué pour préciser leur type afin de pouvoir utiliser ces données lors de la modélisation. Le tableau 17 reporte la fiabilité des liens.

Tableau 17 Volume de liens confirmés et hypothétiques

Solidite du lien	Nombre de liens
Confirme	2239
Non Confirme	118
Total general	2357

Les liens confirmés représentent 95,03% des relations (contre 4,97% pour les liens non confirmés), ce qui donne au graphe relationnel global un bon niveau de fiabilité³⁰³.

Le tableau 18 propose un exemple de lien et de ses attributs et indique le taux de remplissage pour chacun d'entre eux.

³⁰² 12,90% des liens sont de type « autre ».

³⁰³ Cela n'implique pas que graphe corresponde parfaitement à la réalité, mais qu'il est possible de travailler sur les liens matérialisés avec un bon niveau de confiance. Concernant les liens non matérialisés, rien n'indique qu'ils n'existent pas en réalité.

Tableau 18 Colonnes concernant les liens (communes à tous les onglets), exemple de données et taux de remplissage

Colonnes	LIBELLE LIEN	TYPE LIEN	SOLIDITE LIEN	LARGEUR LIEN	DATE LIEN	HEURE LIEN	REF PV	UPVA
Exemple	Conducteur	Utilisateur	Confirmé	3	22/10/2016	14:27:00	0176933/DDJA	40974/01769/2015-33/DDJA
Taux de remplissage	100%	100%	100%	99,87%	40,98%	34,10%	99,92%	99,92%

Ainsi, le taux de remplissage des caractéristiques des liens (84,35%), est extrêmement supérieur à celui des attributs des entités. Il sera ainsi possible d'intégrer des attributs de liens à la modélisation lorsqu'ils s'avèrent pertinents. Les données sur le type et la solidité des liens pourront notamment être prises en compte. Les informations de date et d'heure pourraient s'avérer très utiles pour mettre en évidence des liens qui n'apparaissent pas directement dans les matrices ou, au contraire, montrer que deux entités apparemment liées *via* un intermédiaire sont en réalité temporellement éloignées. Cependant, ces deux attributs ont un taux de remplissage trop faible pour permettre leur intégration au modèle d'analyse : les dates et heures sont manquantes respectivement pour 59,02% et 65,90% des liens, soit plus de la moitié d'entre eux. Cela peut s'expliquer soit par le fait que l'information n'est pas disponible dans la pièce de procédure, soit parce que l'analyste ne l'a pas jugée pertinente.

III.3 Saisie des données : choix du matériel et de la méthode de saisie

Le matériel de travail se constitue initialement de l'ensemble des pièces de procédure en version papier originale. La première phase d'exploration a été compliquée, par manque d'expérience et de méthode. Le matériel de travail, dans sa constitution, était jusque-là inconnu, et il a été nécessaire de se familiariser avec une grande quantité d'information non numérisée ni structurée.

Aussi, durant les trois premiers mois, une méthode d'acquisition pertinente et efficace a été recherchée. Le manque d'expérience pratique a donné lieu à plusieurs tentatives de saisie et structuration des données, qui se sont toutes révélées, *in fine*, inefficaces et/ou inadaptées.

Diverses raisons justifient cela :

- Compte tenu du volume considérable de documents, plus de 20 kg de papier recto-verso, le premier choix a été de traiter les informations parallèlement à la lecture et découverte des pièces de procédure, sans avoir de vue d'ensemble du dossier.
- Les connaissances en analyse criminelle opérationnelle, acquises au cours du parcours étudiant, bien que solides, étaient jusque-alors essentiellement théoriques, ou portaient sur l'exploitation de données préalablement structurées par un tiers.
- De la structuration initiale des données dépendent la finesse, la pertinence et la qualité de l'analyse qui s'ensuit, aussi il n'était pas envisageable de proposer une structure approximative.

Ces raisons nous ont conduit à utiliser la structure existante en gendarmerie pour les analyses en criminalité organisée.

L'un des objectifs du présent de travail était de rechercher si, dans l'ensemble des données à disposition, certaines pouvaient être traitées et analysées autrement, notamment d'un point de vue sociologique, pour apporter une analyse du réseau complémentaire de l'analyse formelle effectuée en gendarmerie. Or, chaque analyste a un raisonnement et une méthode qui lui sont propres. Fort de son expérience, l'analyste qui a saisi les données a réalisé en amont une pré-sélection des informations qu'il considérait ou non comme pertinentes et/ou exploitables avec les méthodes d'analyse utilisées par la suite. En l'occurrence, l'analyste ayant effectué la saisie était en mission à l'étranger, et il n'était pas possible de le contacter pour échanger avec lui et prendre en considération son raisonnement. Reprendre le travail de quelqu'un d'autre sans documentations ou lignes directives sur la méthode et l'expérience personnelles adoptées pouvait s'avérer dangereux, notamment car le risque de ne pas maîtriser les biais induits n'est pas négligeable.

C'est pourquoi il a été décidé de continuer à travailler sur les pièces de procédure en version papier, et de saisir manuellement les informations qu'elles contenaient dans la matrice d'analyse de la Gendarmerie. Le fait de recommencer la saisie permettait de contourner le biais d'expérience et de méthode, propre à chaque individu, évoqué plus haut. L'utilisation de la matrice vierge comme structure des données permettait en revanche d'avoir un matériel de base similaire à celui actuellement traité en gendarmerie, sans biais de sélection.

Cependant, comme évoqué plus haut, le volume de données à saisir était colossal ; il s'est révélé illusoire d'imaginer effectuer trois années de travail d'un analyste en seulement quelques mois, surtout sans pré-sélection des données, c'est-à-dire en saisissant, sans le filtre de l'expérience, l'ensemble des données contenues dans le dossier. Ainsi, pour les raisons

mentionnées dans la section « Aspects pratiques », une nouvelle stratégie a été adoptée en décembre 2020 : celle de finalement travailler sur les matrices pré-saisies par l'analyste, tout en ayant conscience de l'incomplétude potentielle des informations enregistrées³⁰⁴, et des biais induits par la méconnaissance de la méthode de saisie. Les données brutes – ou initiales – de cette thèse sont donc des données secondaires, issues d'une première sélection sur l'ensemble des pièces de procédure.

Les matrices de données originales, c'est-à-dire non analysées, remplies par l'analyste de la SR Paris XX^{ème} ont été obtenues fin février 2021, et constituent finalement le matériel de travail adopté pour la suite.

III.4 Principaux outils de travail

a) L'outil *Stemic* développé et commercialisé par Intactile Design

Développé par la société Intactile Design depuis près de 15 ans, *Stemic*³⁰⁵, est un support de réflexion basé sur des modèles de connaissances pour l'aide à la décision. Cet environnement s'avère particulièrement adapté aux domaines de l'enquête et du renseignement criminels, bien que sa portée soit bien plus large.

En effet, les deux domaines susmentionnés nécessitent d'être en mesure de poser, visualiser ou cartographier ses idées, les cheminements réflexifs, ainsi que les éléments à disposition. Notamment, en représentant visuellement les informations à disposition et les connexions les unissant, l'outil permet d'effectuer une sélection pertinente des aspects sur lesquels se concentrer, de garder en mémoire les raisonnements effectués, ce qui est essentiel : la traçabilité non seulement de la donnée, mais aussi des connaissances, informations et renseignements³⁰⁶ extraits est cruciale.

Stemic a été utilisé aussi bien pour illustrer les raisonnements employés et développés au cours de la thèse que pour se repérer dans l'enquête. En outre, l'outil étant toujours en développement, nous avons pu faire part de retours utilisateurs, en identifiant et en anticipant

³⁰⁴ Nous ne remettons pas ici en cause les compétences de l'analyste, dont la pré-sélection est pertinente et efficace dans le temps imparti par l'enquête et pour répondre à ses objectifs, mais cherchons à souligner le fait qu'il aurait été plus rigoureux, non seulement en termes de recherches, mais aussi de renseignement criminel, de travailler sur l'ensemble des informations contenues dans le dossier dont nous disposons.

³⁰⁵ <https://intactile.com/projets/Stemic/>

³⁰⁶ Termes utilisés au sens du cycle du renseignement (voir chapitre I, section I.4).

les besoins – tels que de nouvelles fonctionnalités – des usagers du domaine de l'enquête et du renseignement criminels auxquels ils se destine notamment.

b) L'outil *Tulip* développé par le LaBRI

Pour la visualisation et l'analyse relationnelle dans son ensemble des données massives dont nous disposons, l'outil *Tulip*³⁰⁷, développé par le laboratoire bordelais de recherche en informatique (LaBRI – Université de Bordeaux) a été utilisé. *Tulip* est une solution logicielle en source ouverte dédiée à la visualisation et l'analyse de grands graphes. Ce logiciel permet, à l'aide d'une interface de programmation en langage python, d'écrire des algorithmes pour effectuer des traitements, des analyses et des calculs spécifiques au besoin n'étant pas initialement disponibles.

Comme nous le verrons par la suite (*cf.* chapitre IV), il est impératif de conserver le lien avec l'information originale tout au long des traitements appliqués aux données. La plateforme *Tulip* permet de garder en mémoire un historique complet des opérations effectuées, elle est donc parfaitement adaptée au cas d'usage. En effet, *Tulip* gère une hiérarchie de graphes et des collections de propriétés associées aux éléments du graphe à la manière du mécanisme d'héritage des langages de programmation à objets³⁰⁸. Ainsi, un sommet présent dans un sous-graphe n'est pas dupliqué en mémoire, c'est-à-dire qu'il n'est pas stocké deux fois : il s'agit du même objet – d'un point de vue informatique – que dans son graphe parent. Les propriétés associées aux éléments du sous-graphe sont héritées du graphe parent, mais peuvent être masquées par une nouvelle propriété locale au sous-graphe. Par exemple, dans la figure 18, la propriété *color2* dans le sous-graphe 1.2 masque la propriété *color1* de la racine de la hiérarchie. Les sommets du sous-graphe 1.2 héritent de leurs propriétés initiales dans le graphe parent, propriétés qui peuvent ensuite être modifiées localement : un sommet du sous-graphe 1.2 peut donc avoir une couleur différente de celle qu'ils avaient dans le graphe racine. Lorsqu'une propriété d'un sommet ou d'une arête d'un sous-graphe est modifiée, cela ne masque en revanche pas la propriété initiale dans le graphe parent.

³⁰⁷ <https://Tulip.labri.fr/site/>

³⁰⁸ David Auber, Daniel Archambault, Romain Bourqui, Maylis Delest, Jonathan Dubois, Antoine Lambert, Patrick Mary, Morgan Mathiaut, Guy Melançon, Bruno Pinaud, Benjamin Renoust et Jason Vallet, « TULIP 5 », dans Reda Alhajj et Jon Rokne (dir.), *Encyclopedia of Social Network Analysis and Mining*, New York, NY, Springer New York, 2017, pp. 1-28.

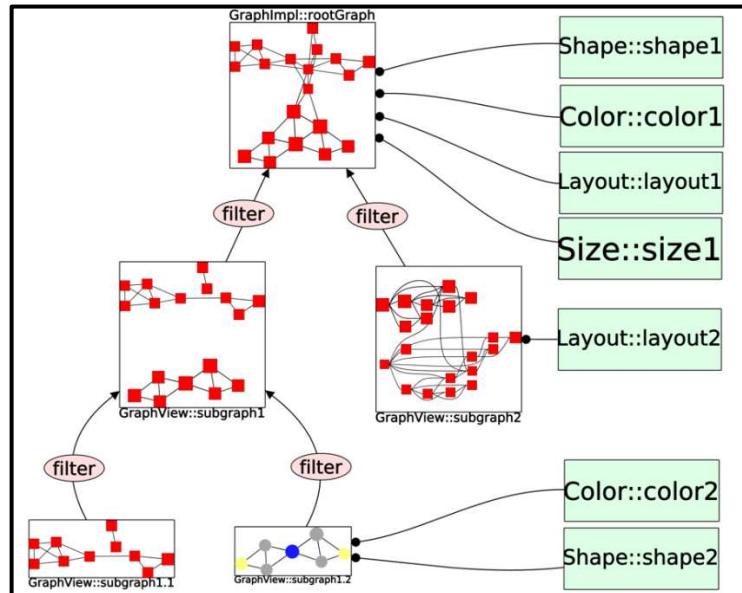


Figure 18 Modèle de la hiérarchie de graphes de *Tulip*³⁰⁹

Tulip a ainsi été utilisé dans un premier temps lors des traitements des données puis, dans un second temps, pour l'analyse du réseau social que constitue l'organisation criminelle étudiée, la plateforme étant dotée de nombreux algorithmes et fonctions permettant d'effectuer des calculs et opérations en ARS, tels que des mesures de densité ou de centralité.

En la personne de Bruno Pinaud, MCF et HDR en informatique, un partenariat de recherche a été développé avec le LaBRI, dont les connaissances en théorie des graphes, réflexions en analyse de réseaux sociaux et compétences en langage python ont été non seulement d'un apport déterminant, mais aussi à la source de nouvelles idées. Ainsi, la transformation de modèle (voir section V.1) a été réalisée conjointement avec le LaBRI.

c) Traitements de données : *Excel* et *RStudio*

L'outil *Excel* a été employé pour les pré-traitements et traitements de données sous leur forme matricielle. Pour les traitements et analyses statistiques plus avancés des données, l'interface de programmation en langage *R* *RStudio*³¹⁰, disponible en source ouverte, a été utilisée. Le langage *R* est un langage dédié à la manipulation statistique.

³⁰⁹ Source : *Ibid.* avec l'aimable autorisation des auteurs.

³¹⁰ <https://www.RStudio.com/>

CHAPITRE IV DU PRÉ-TRAITEMENT DES DONNÉES BRUTES AU GRAPHE MULTIVARIÉ

La méthodologie générale employée s'inscrit dans le cycle du renseignement présenté au chapitre I et schématisé dans la figure 19.

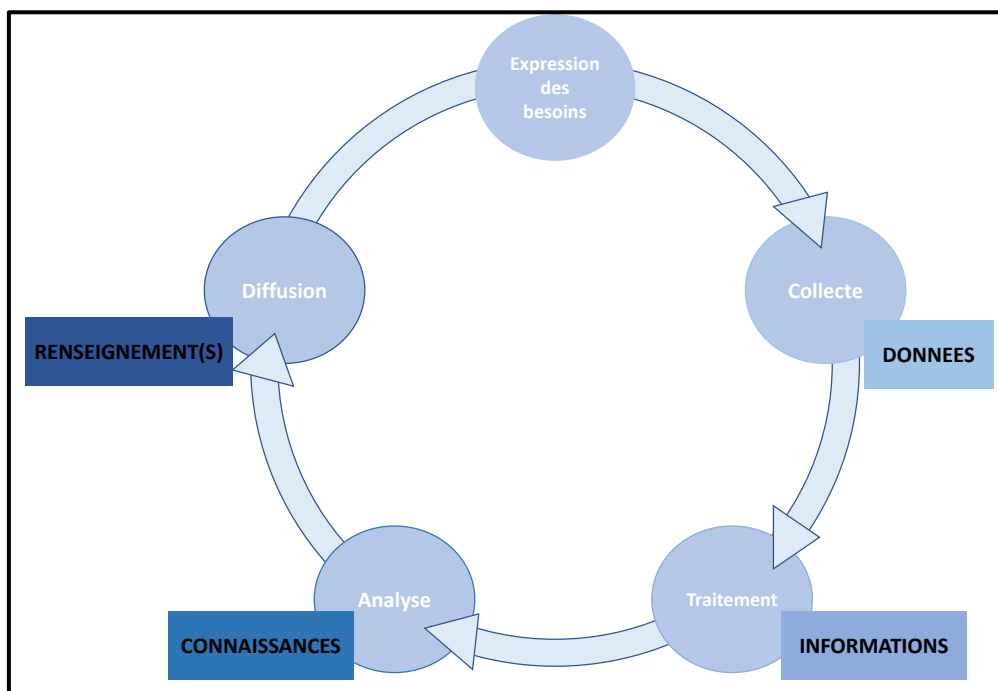


Figure 19 Le cycle du renseignement

Le cycle du renseignement (figure 19) est un outil méthodologique formalisant la création de renseignement(s) à partir de données brutes en appliquant les étapes de compréhension du besoin, de collecte de la donnée adéquate, de traitement de l'information, d'analyse de la connaissance et enfin de diffusion du renseignement.

L'expression (et la compréhension) du besoin est le préambule du cycle du renseignement : elle est essentielle pour orienter, coordonner et organiser la recherche et le recueil des données afin de répondre au mieux à la demande. Dans le cadre de la présente thèse, le besoin est exprimé dans l'objectif de recherche : proposer une méthode de détection des fonctions et positions stratégiques dans un réseau criminel afin d'améliorer la compréhension et la connaissance par les forces de l'ordre de la structure d'organisations criminelles. Or, dans notre cas, la donnée initiale répond à un besoin distinct : celui de l'enquête judiciaire, c'est-à-dire l'établissement de faits, la recherche (individualisation et localisation) d'auteurs et la

collecte d'éléments probants (qu'ils soient à charge ou à décharge)³¹¹. De plus, sa collecte est soumise au cadre juridique spécifique de l'investigation, à la différence du recueil de données dans le cadre du renseignement criminel, qui ne dispose pas d'un cadre juridique explicitement prévu dans les textes³¹² (figure 20).

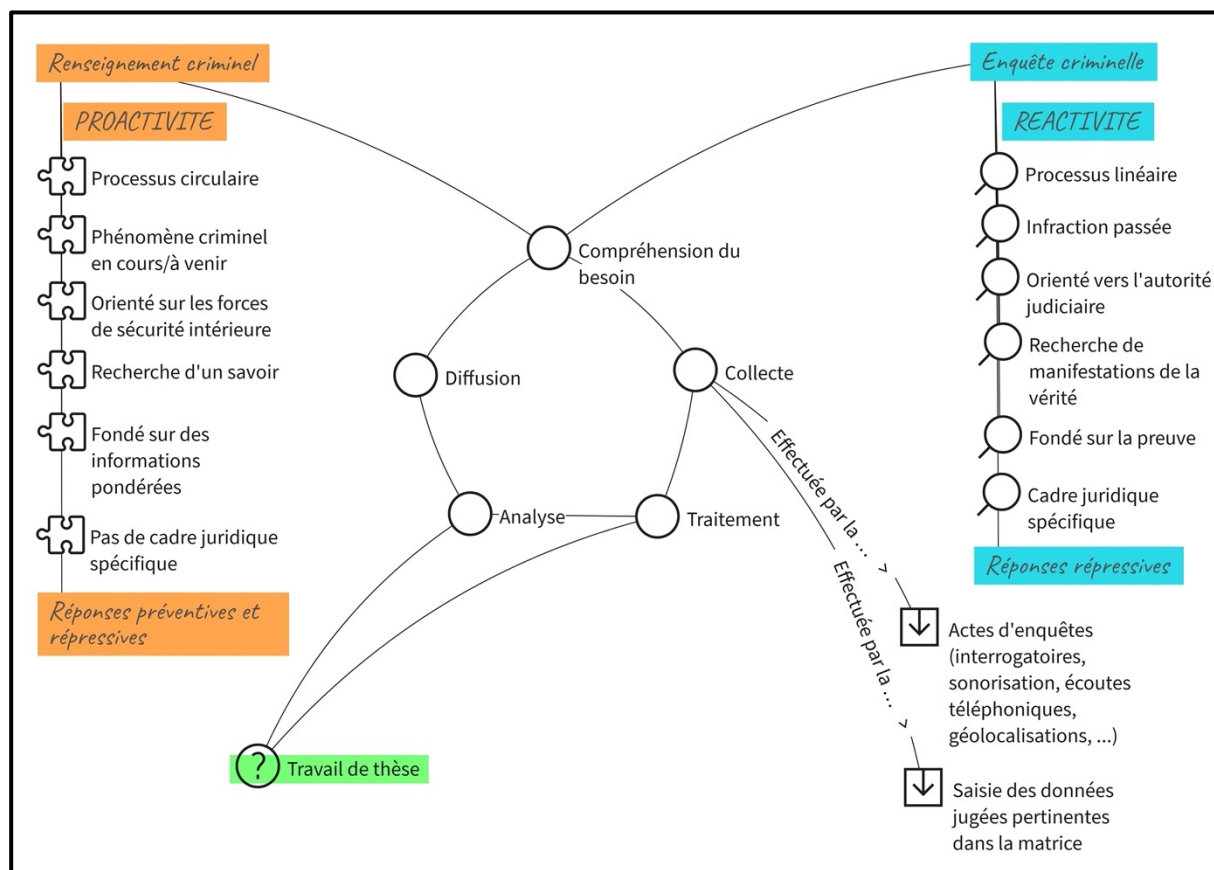


Figure 20 Une donnée, deux besoins distincts.

Ce changement de paradigme, intimidant de premier abord, est pourtant partie intégrante du travail de l'analyste en renseignement : le renseignement criminel se fonde sur la réalité de la délinquance constatée par les unités opérationnelles, et donc, par essence, sur les données collectées et produites par l'enquête judiciaire (procès-verbaux de procédures, interrogatoires, débriefing d'enquêtes...). Afin d'en extraire des connaissances de nature différente (analyse de risques, détection de points chauds, analyse de modes opératoires, de profil d'auteurs et de victimes, détection de *patterns*, compréhension de mécanismes...) de celles recherchées durant l'investigation, l'étape de traitement de la donnée est fondamentale : de sa qualité et de sa

³¹¹ Clément De Maillard, « La France et le renseignement criminel : entre volonté et réalité, une ambition à écrire », *Sécurité et stratégie*, vol. 17, n° 2, 2014, pp. 49-59.

³¹² Le renseignement criminel se trouve à l'intersection du Code de Sécurité intérieure encadrant le renseignement de manière générale et du Code de Procédure pénale qui encadre l'enquête et l'information judiciaire, l'échange d'information entre les services et certains traitements de données à caractère personnel. Jérôme Barlatier, cours « Cadre juridique du Renseignement criminel », stage d'analyse en renseignement criminel, SCRC, 2018.

minutie dépendront les possibilités d'exploitation. En effet, la donnée doit souvent être (re)formatée et préparée en regard du nouvel objectif – produire du renseignement criminel – dans la mesure où la structure initiale peut s'avérer incompatible avec les analyses et mesures que l'analyste en renseignement criminel souhaite effectuer. Si la donnée a été collectée et saisie pour répondre à un besoin d'investigation, elle a aussi et surtout été stockée et structurée pour être traitée avec les méthodes et outils d'enquête, souvent différents des méthodes d'analyse en renseignement criminel³¹³.

Ce changement de paradigme est matérialisé lors de la transition de l'objectif de recherche à la problématique. En effet, cette transition formalise le fait que cette thèse se concentre sur des données issues d'actes d'enquête et d'analyse criminelles : peut-on combiner les outils analytiques issus de multiples disciplines pour développer, *sur la base de données telles que collectées et saisies lors des enquêtes de gendarmerie*, une méthodologie de représentation, de modélisation et d'analyse de réseaux criminels aboutissant à une détection des fonctions et positions stratégiques pour le fonctionnement de l'organisation ?

IV.1 Préambules de l'analyse : pré-traitements et traitements des données

Les données brutes – ou initiales – sont ici constituées de la matrice d'analyse décrite au chapitre précédent et renseignée par l'analyste en charge du dossier K**** sur la base des éléments de procédure d'investigation.

L'objectif des analyses est d'en extraire des connaissances pour répondre au besoin d'une meilleure compréhension des structures et fonctionnements de réseaux criminels : il s'agit de structurer la donnée en vue de son exploitation par le renseignement.

La méthode de traitement appliquée par les analystes en renseignement criminel³¹⁴ se décompose en cinq principes : classer les données (stocker, recenser, tracer), les uniformiser

³¹³ Il est important de préciser que cette difficulté découle directement du choix méthodologique de travailler sur la donnée secondaire plutôt que primaire à savoir sur la matrice d'analyse saisie dans le cadre de l'investigation plutôt que sur les procès-verbaux. La donnée primaire est le matériel brut récolté, tel que le constat qu'un individu se trouve à un endroit donné à un moment donné ; la donnée secondaire est la donnée saisie dans une matrice d'analyse, mais pas encore traitée. Elle deviendra une information lorsqu'elle aura été formatée pour l'analyse dans le temps de thèse imparti, une saisie des données contenues dans les PV dans une matrice vierge par la doctorante aurait permis de les stocker avec une structure et un format directement compatibles avec la méthode d'analyse proposée dans la suite du manuscrit.

³¹⁴ Jérôme Barlatier, cours « Traitement de l'information criminelle », *op. cit.*

(nettoyer, rendre comparable), les comprendre (indexer, reconnaître et fusionner les entités), les ordonner (structurer, établir des ontologies, effectuer les premiers rapprochements) et enfin visualiser les données, que ce soit sous forme de tableurs, de bases de données, de cartographies, de représentations temporelles, de schémas relationnels, cartes heuristiques, représentations statistiques...

Ainsi que cela a été détaillé au chapitre précédent, la collecte des données a été réalisée par les enquêteurs en charge de l'affaire, et leur saisie dans la matrice d'analyse (stockage, recensement, traçabilité) a été effectuée par l'analyste criminel de la SR de Paris.

Le travail de traitement des données débute donc au moment de la prise en main de la matrice par l'étape d'uniformisation.

a) Uniformisation structurelle des données, préambule nécessaire à l'extraction automatisée de connaissance

Il est important de préciser en avant-propos que tous les traitements et corrections ont été effectués de manière versionnée sur une copie des données initiales et que les fichiers originaux ont été conservés en l'état.

La toute première étape a consisté à réunir les différents onglets de la matrice en un seul – évoqué dans la suite du chapitre comme le « fichier unique », afin d'uniformiser les données, permettant ainsi non seulement de mieux les intégrer d'un point de vue cognitif, mais aussi de se focaliser sur les attributs exploitables du point de vue du renseignement criminel et ainsi de rendre l'import des données dans le logiciel *Tulip* uniforme.

Initialement, chacun des six onglets de la matrice exploités lors de l'investigation correspondait au type d'entité depuis lequel les relations référencées ont été établies. Les six onglets, « personnes », « personnes morales », « télécom », « lieux », « objets », et « véhicules » contenaient des attributs relatifs aux entités sources, donc de natures différentes (date de naissance pour les personnes, marque pour les véhicules, opérateurs pour les téléphones...). Outre les identifiants et les icônes décrivant les entités, compte tenu du faible taux de remplissage de la majorité des attributs, seules les deux propriétés les plus renseignées et les plus pertinentes ont été conservées pour la suite des traitements. La figure 21 illustre les attributs retenus pour chaque type d'entité.

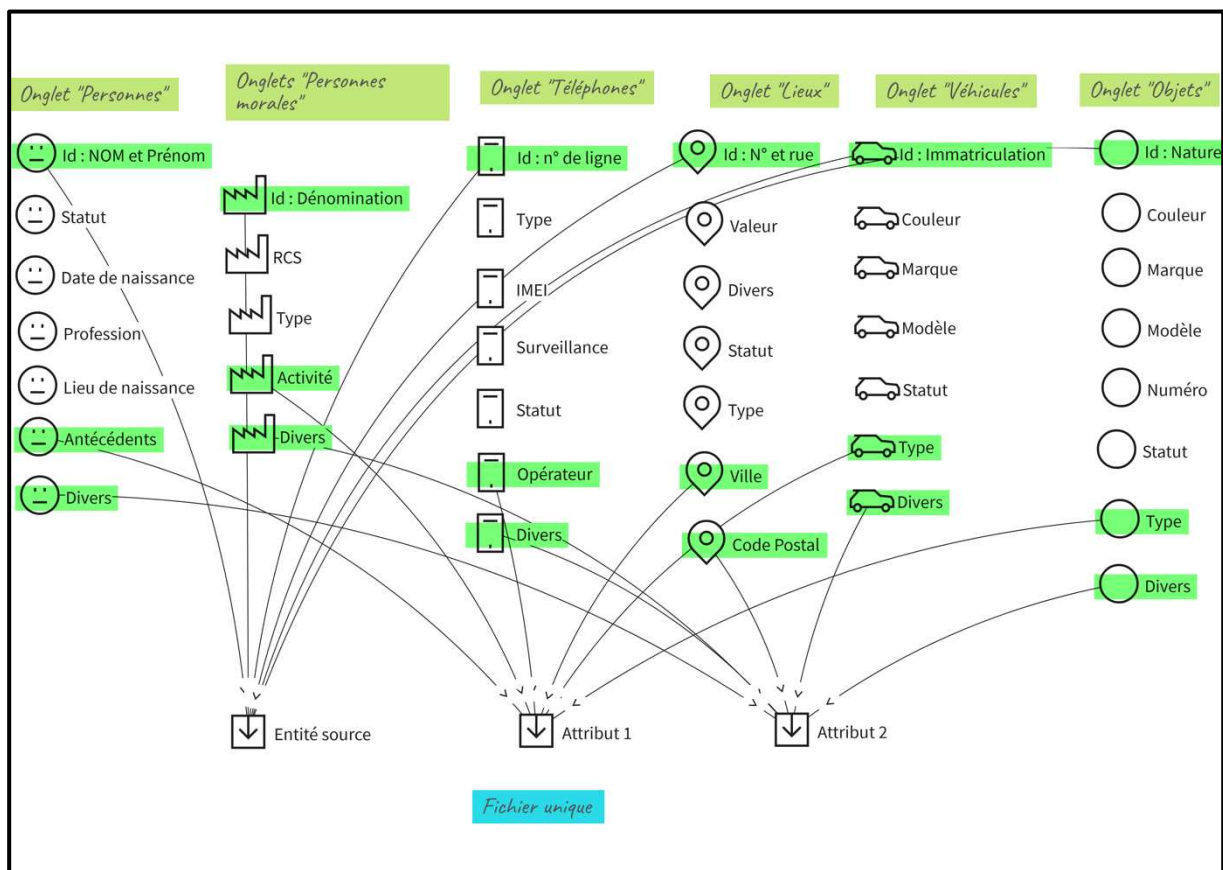


Figure 21 Sélection des attributs conservés³¹⁵

Les attributs des liens étant identiques dans tous les onglets, ils restent inchangés. À ce stade, le fichier unique contenant toutes les relations saisies par l'analyste sous forme d'*edge list* est structuré comme illustré dans le tableau 19.

Tableau 19 Structure et exemple³¹⁶ de données après réunion des onglets.

ENTITE SOURCE	ICONE SOURCE	Attribut 1	Attribut 2	LIBELLE LIEN	TYPE LIEN	SOLIDITE	DATE LIEN	HEURE LIEN	REF PV	UPVA	DESTINATION	ICONE DESTINATION
DEVAUT Léon	Homme	Vol par effraction, enlèvement, séquestration	Fausse identité	Récupère un sac noir et blanc	Autre	Confirmé	13/08/2016	12:27	0146828 /XX	30YY/01 468/201 6-21Z/XX	AB-123-CD	Voiture

Les données ont alors été indexées avec l'attribution d'un identifiant unique à chaque entité, ce qui permettra par la suite de pseudonymiser les données (tableau 20) – c'est-à-dire de travailler avec un matériel non nominatif³¹⁷, tout en étant capable de remonter aux données brutes si nécessaire *via* un fichier de correspondance entre les données nominatives et les identifiants – puis d'anonymiser les données après la thèse en supprimant le fichier de correspondance.

³¹⁵ Schéma réalisé avec l'outil *Stemic*, développé par la société Intactile Design.

³¹⁶ Données anonymisées : pour des raisons de confidentialité, seules des informations fictives sont ici renseignées.

³¹⁷ Comme convenu lors des accords de cession de données signés avec la Gendarmerie nationale.

Tableau 20 Création d'identifiants uniques pour chaque entité

ENTITE SOURCE	ID	ICONE SOURCE	Attribut 1	Attribut 2	LIBELLE LIEN	TYPE LIEN	SOLIDITE	DATE LIEN	HEURE LIEN	REF PV	UPVA	DESTINATION	ID	ICONE DESTINATION
DEVAUT Léon	57	Homme	Vol par effraction, enlèvement, séquestration	Fausse identité	Récupère un sac noir et blanc	Autre	Confirmé	13/08/2016	12:27	01468 28/XX	30YY/ 01468 /2016 21Z/X X	AB-123-CD	1684	Voiture

b) Nettoyage et appropriation du contenu

Une fois la structure harmonisée, les données doivent être nettoyées. Les processus liminaires de nettoyage de données ont été appliqués : remplacement des caractères spéciaux, suppression des doublons, gestion de la casse (passage des entités et de leurs attributs en majuscules, des liens et de leurs attributs en minuscule). Ces premiers traitements réalisés, le fichier unique a été minutieusement exploré afin de mieux se familiariser avec son contenu et de le comprendre : étude de la typologie des relations, des types d'entités reliées, des libellés des liens... Il est apparu que des corrections du contenu devaient être appliquées. En effet, si celui-ci était parfaitement compréhensible sémantiquement parlant et donc exploitable par une intelligence humaine, les données n'étaient pas assez homogènes pour être exploitables de manière automatisée par des algorithmes. De fait, les corrections nécessaires n'étaient pas non plus automatisables, et ces dernières ont été effectuées manuellement en parcourant une à une les 2 396 lignes, cellule par cellule, du fichier unique. Dans un premier temps, les erreurs de frappe et d'orthographe ont été corrigées. Dans un second temps, pour un certain nombre d'entités (en particulier les personnes et les personnes morales), des orthographes multiples du label (nom et prénom pour les personnes et dénomination pour les personnes morales) conduisant à l'existence d'entités multiples pour un seul et même élément ont été détectées. Les labels de ces entités ont donc été uniformisés afin de n'avoir qu'un unique identifiant et un unique nœud par entité dans le graphe de visualisation qui sera généré par la suite (tableau 21).

Tableau 21 Homogénéisation des labels des entités³¹⁸

Orthographes multiples	Identifiant	Orthographe unique	Identifiant
THIBAUT ROSIER	1	THIBAUT ROSIER	1
THIBAUT ROSIET	42	THIBAUT ROSIER	1
THIBAUT ROSIER	1103	THIBAUT ROSIER	1
LA QUINCAILLERIE DU PORT	38	LA QUINCAILLERIE DU PORT	38
LA QINCAILLEIRE DU PORT	394	LA QUINCAILLERIE DU PORT	38

³¹⁸ Données nominatives fictives.

La troisième étape de correction manuelle a concerné l'harmonisation entre les entités sources et destinations d'une part et les liens les unissant d'autre part. En effet, dans l'optique d'automatiser une partie des traitements d'ordre linguistique qui seront appliqués aux données lors de la phase d'analyse³¹⁹, il est nécessaire que chaque ligne du fichier soit sémantiquement cohérente. Les tableaux 22 et 23 illustrent ces corrections.

Tableau 22 Exemple d'anomalie sémantique

Entité source	Libellé lien	Entité destination	Cohérence sémantique
VÉHICULE A	Discute avec	VÉHICULE B	Non

Tableau 23 Exemple de standardisation des données

Entité source	Libellé lien	Entité destination	Cohérence sémantique
INDIVIDU A	Conduit	VÉHICULE A	Oui
INDIVIDU B	Conduit	VÉHICULE B	Oui
INDIVIDU A	Discute avec	INDIVIDU B	Oui

La quatrième étape d'uniformisation a consisté à travailler sur l'harmonisation des typologies d'entités. En effet, comme abordé dans la section III.2 « Description des données à disposition : forme, structure et contenu », la typologie des entités présentes dans les données est matérialisée dans la colonne « icône » (tableau 24), ce qui facilite la représentation des entités dans les visualisations graphiques.

Tableau 24 Liste initiale des différentes entités (information stockée dans la colonne « icônes »)

Compte
Restaurant
Café
Magasin
Lieu
Appartement
Aéroport
Maison
Garage
Document

Carte de crédit bancaire (dos)
Drogue
Homme
Femme
Anonyme
Garçon
Fille
Clan
Bureau
Carte SIM

Téléphone
Téléphone portable
Courrier électronique
Serveur
Voiture
Moto
4X4
Fourgon
Camion-plateau

Or, le parcours des lignes du fichier lors des étapes précédentes a également révélé que certains types d'entités pouvaient être associés à plus d'une icône et réciproquement. Ces typologies multiples ne concernaient que les personnes morales. À titre d'exemple, les enseignes de *fast-food* étaient référencées tantôt comme un restaurant, un magasin, un bureau

³¹⁹ Cf. chapitre V, section V.1 « Extraction sémantique du réseau social depuis le graphe multivarié : approche par la science des réseaux ».

ou un lieu. La même anomalie pouvait en outre être observée d'une ligne à l'autre pour une même enseigne. À l'inverse, l'icône « magasin » pouvait aussi bien désigner une boutique qu'un bar ou un restaurant. La nomenclature suivante a donc été mise en place (tableau 25).

Tableau 25 Nomenclature établie pour référencer les entités par icônes

Icône	Type de personne morale
Bureau	Bureaux, locaux et sièges sociaux d'entreprises (n'accueillant pas de public)
Café	Bars et cafés
Lieu	/
Magasin	Boutiques, supermarchés, centres commerciaux
Restaurant	Restaurants et <i>fast-food</i>

Enfin, la cinquième et dernière étape a consisté à standardiser l'utilisation de la colonne « Solidité lien » qui indique pour chaque ligne si la relation entre les entités source et destination est confirmée ou non. Si cette colonne a été utilisée lors de la saisie des données dans la matrice pour caractériser certaines relations comme hypothétiques, cela n'était pas systématique. En effet, pour certains liens présumés, le caractère hypothétique de la relation était directement stocké dans la colonne « Libellé lien », et non dans la colonne « Solidité lien », comme illustré dans le tableau 26.

Tableau 26 Stockage initial de certains liens présumés

ENTITE SOURCE	LIBELLE LIEN	SOLIDITE LIEN	ENTITE DESTINATION
INDIVIDU A	Susceptible d'avoir récupéré de l'argent	Confirmé	BOUTIQUE X
INDIVIDU B	Semble travailler avec	Confirmé	INDIVIDU C
INDIVIDU D	Est en contact (transaction supposée)	Confirmé	INDIVIDU E
INDIVIDU F	Est peut-être le conducteur	Confirmé	VÉHICULE Y

Pour toutes les lignes concernées, le caractère incertain a été retiré du libellé du lien et explicité par l'attribut de solidité, comme le montre le tableau 27. À défaut d'uniformité sémantique, ces corrections n'ont pu être automatisées et ont dû être effectuées manuellement.

Tableau 27 Stockage après standardisation des liens présumés

ENTITE SOURCE	LIBELLE LIEN	SOLIDITE LIEN	ENTITE DESTINATION
INDIVIDU A	A récupéré de l'argent	Non confirmé	BOUTIQUE X
INDIVIDU B	Travaille avec	Non confirmé	INDIVIDU C
INDIVIDU D	Est en contact (transaction)	Non confirmé	INDIVIDU E
INDIVIDU F	Est le conducteur	Non confirmé	VÉHICULE Y

Ainsi, après plusieurs mois de pré-traitement manuel ligne par ligne et cellule par cellule, les données sont structurées et uniformisées : les traitements et analyses peuvent débiter.

c) Standardisation du contenu par la création d'ontologies sémantiques :
préparation aux traitements automatisés

Les données prétraitées contiennent 1 147 entités distinctes réparties en 31 types (colonne « icône ») liées par 2 357 relations, référencées sous 513 libellés différents. Le nombre d'entités et de liens est trop élevé pour pouvoir les traiter de manière automatique, il est donc nécessaire de créer des ontologies pour les classifier afin de pouvoir travailler sur des catégories plus larges et moins nombreuses. La première concerne les entités et est représentée dans le tableau 28. Le nombre de classes à gérer diminue ainsi de 31 types à 10 catégories détaillées dans le tableau 29.

Tableau 28 Ontologie des entités

ICONE SOURCE (entité)	CATEGORIE ENTITE
Compte	BANQUE
Café	LIEU PUBLIC
Restaurant	LIEU PUBLIC
Magasin	LIEU PUBLIC
Lieu	LIEU (ADRESSE)
Appartement	LIEU PRIVÉ
Aéroport	LIEU PUBLIC
Ville	LIEU PUBLIC
Maison	LIEU PRIVÉ
Garage	LIEU PUBLIC
Document	OBJET
Carte de crédit bancaire (dos)	OBJET
Drogue	OBJET
Homme	PERSONNE
Femme	PERSONNE
Anonyme	PERSONNE
Garçon	PERSONNE
Fille	PERSONNE
Clan	PERSONNE
Bureau	PERSONNE MORALE
Carte SIM	TÉLÉCOM
Téléphone	TÉLÉCOM
Téléphone portable	TÉLÉCOM
Courrier électronique	TÉLÉCOM
Serveur	TÉLÉCOM
Voiture	VÉHICULE
Moto	VÉHICULE
4x4	VÉHICULE
Fourgon	VÉHICULE
Camion-plateau	VÉHICULE
Parking	LIEU PUBLIC

Tableau 29 Liste des catégories d'entités utilisées dans la suite du travail

CATEGORIES ENTITES
PERSONNE
VÉHICULE
TÉLÉCOM
PERSONNE MORALE
LIEU (ADRESSE)
LIEU PRIVÉ
LIEU PUBLIC
AUTRE
BANQUE
OBJET

Il est important pour la suite du travail de spécifier que les entités référencées par les icônes « Magasin », « Restaurant » et « Café », *i.e.* les locaux professionnels accueillant du

public ont été ici catégorisés en « LIEU PUBLIC », tandis que les entités « Bureau » sont définies comme des « PERSONNES MORALES ». En effet, si ces quatre types font référence à des entités professionnelles, les trois premiers sont essentiellement associés à des liens de localisation, tandis que le dernier est très majoritairement associé à l'entreprise en soi. Ce choix permettra par la suite d'appliquer des traitements automatiques prenant en compte cette distinction existante dans les données.

La seconde ontologie répartit les 518 libellés de liens en 19 catégories. Cette classification est effectuée sur la base de la sémantique des différents libellés de lien. En informatique, on parle de « concepts » (qui correspondent à nos catégories), décrits par des « termes », à savoir l'ensemble des synonymes et expressions utilisés pour parler d'un concept donné (ici, les libellés)³²⁰. Cette ontologie cherche à rendre compréhensible par des algorithmes la signification des liens, en se rapprochant du fonctionnement cognitif d'une intelligence humaine. L'ontologie complète est disponible en annexe 3; les tableaux 30 et 31 représentent respectivement un extrait de l'ontologie et la liste des 19 catégories de liens.

Tableau 30 Extrait de l'ontologie de classification sémantique des liens

LIBELLE LIEN	CATEGORIE LIEN
Discute	En contact
Récupère 700000€	Échange objet ou argent
Conducteur	Utilisateur
Entre avec un sac en papier blanc	Échange objet ou argent
Est titulaire de l'abonnement EDF du 23/01/2014 au 13/09/2016	Adresse
Se trouve	Localisation
Est arrêté par la police au volant	Utilisateur
Monte à bord après avoir chargé deux sacs	Utilisateur
Quitte l'aéroport de ROISSY à bord	Utilisateur
Entre en contact	En contact
Est interpellé par la sûreté territoriale des hauts de seine avec	Antécédents
Utilise	Utilisateur
Mariée en 2013	Familial/amical
Demeure (passeport 02/03/2015)	Adresse
Envoie mandat cash	Échange objet ou argent
Est gérant de l'épicerie	Professionnel
Est propriétaire (n'est pas sur le PV)	Titulaire
Identifié comme	Est
Vendeur appartement	Autre
Mis en cause (2005)	M.E.C.
N'est pas le père de	Non lien

³²⁰ Dans un souci de clarté, nous utiliserons, par abus de langage, la terminologie des analystes plutôt que celle issue de l'informatique.

Tableau 31 Liste des catégories de liens utilisées dans la suite du travail

VALEURS UNIQUES LIENS
En contact
Échange objet ou argent
Utilisateur
Adresse
Localisation
Titulaire
Antécédents
Familial/amical
Professionnel
Est
Autre
M.E.C.
Criminel
Même nom
Fait partie
Lien interdit
Non lien
Lien
Même adresse

IV.2 Visualisations graphiques

a) Création du graphe relationnel *n-mode*

Les visualisations ont été effectuées avec le logiciel *Tulip* développé par le LaBRI. L'ensemble des scripts et algorithmes d'import et de traitement ont été conçus et paramétrés conjointement par la doctorante et Bruno Pinaud³²¹, maître de Conférences HDR au LaBRI et intégralement écrits par Monsieur Pinaud.

La première étape consiste en l'import des données dans l'outil *Tulip*. Le fichier d'entrée est le fichier unique, sous forme « d'*edge list* », nettoyé et standardisé obtenu suite aux prétraitements décrits dans la section précédente. L'import a été fait *via* l'interface python de *Tulip* avec un script dédié afin de spécifier les paramètres d'import. *Tulip* gère les données

³²¹ <https://www.labri.fr/perso/bpinaud>.

relationnelles en important dans un premier temps les sommets puis en ajoutant les arêtes correspondantes.

Concernant les entités, conformément aux accords de confidentialité signés avec le SCRC en vertu de la circulaire ministérielle n°30053 du 3 septembre 2019 relative à l'export de données à destination de la recherche scientifique, seules ont été importées les colonnes « Catégorie entité », « Identifiant » et « Icône source » (afin de faciliter la représentation visuelle des différentes entités). Ainsi, aucune donnée nominative ou personnelle n'a été importée : les nœuds sont référencés (labellisés) à l'aide de l'identifiant unique qui leur a été attribué à des fins de pseudonymisation durant les phases d'analyse puis d'anonymisation lorsque les travaux seront terminés. Il est donc toujours possible pour la doctorante d'accéder aux identités réelles et à leurs attributs par le biais du fichier de correspondance entre les entités et leurs identifiants.

Concernant les liens, seules les colonnes « Catégorie liens » et « Solidité lien » ont été importées. En effet, comme exposé plus haut, les libellés sont trop nombreux pour être exploitables par les traitements mis en place par la suite. En outre, la description des données dans le chapitre III montre que malheureusement, les informations de date et d'heure sont au contraire trop peu nombreuses pour pouvoir être prises en compte dans le modèle. Cependant, lorsque la donnée est disponible et si besoin, il sera toujours possible d'y accéder dans le fichier d'entrée qui reste inchangé. Enfin, des colonnes supplémentaires, utilisées au cours des traitements ont été créées³²².

³²² Cf chapitre V sections V.1 b) « Méthodologie de transformation : inspiration issue des concepts de grammaires formelles » et d) « Pré-traitements graphiques : réduction de l'espace de recherche. »

Les tables « *Nodes* » et « *Edges* » se présentent comme indiqué respectivement tableaux 32 et 33.

Tableau 32 15 premières lignes du fichier « *Nodes* »

ID	CATEGORIE
1	PERSONNE
68	PERSONNE
67	PERSONNE
873	LIEU (ADRESSE)
777	VÉHICULE
874	LIEU (ADRESSE)
763	VÉHICULE
90	PERSONNE
776	VÉHICULE
147	PERSONNE
875	PERSONNE
2	PERSONNE
16	PERSONNE
770	VÉHICULE
558	LIEU PUBLIC

Tableau 33 15 premières lignes du fichier « *Edges* »

CATEGORIE	SOLIDITE LIEN
En contact	Confirmé
En contact	Confirmé
Échange objet ou argent	Confirmé
Utilisateur	Confirmé
Échange objet ou argent	Confirmé
Utilisateur	Confirmé
Utilisateur	Confirmé
Utilisateur	Confirmé
En contact	Non Confirmé
Utilisateur	Confirme
Utilisateur	Confirme
Échange objet ou argent	Confirme
Échange objet ou argent	Confirme
Échange objet ou argent	Confirme
Échange objet ou argent	Confirme

Certaines lignes ont nécessité une étape supplémentaire lors de l'import : la création de nouveaux nœuds, appelés méta-nœuds, qui regroupent plusieurs entités équivalentes. En effet, sur l'ensemble du fichier, certaines lignes mettent en évidence de multiples identités pour une seule et même personne, comme illustré dans le tableau 34.

Tableau 34 Mise en évidence (information confirmée) d'identités multiples

ENTITE SOURCE	CATEGORIE ENTITE SOURCE	CATEGORIE LIEN	SOLIDITE LIEN	ENTITE DESTINATION	CATEGORIE ENTITE DESTINATION
MADAME X	PERSONNE	Est	Confirmé	MADAME Y	PERSONNE

Il est indispensable d'utiliser cette information pour l'analyse du réseau. Ce type de lien est reporté pour 4 personnes, ayant chacune deux identités distinctes. Pour chacun de ces quatre cas, nous avons généré un sous-graphe contenant les deux nœuds initiaux réunis dans un méta-nœud, ce qui permet de conserver l'information originale. Ainsi, quatre méta-nœuds viennent s'ajouter aux 1 147 nœuds existant dans les données (figure 22) : soit un total de 1 151 nœuds dans le graphe.

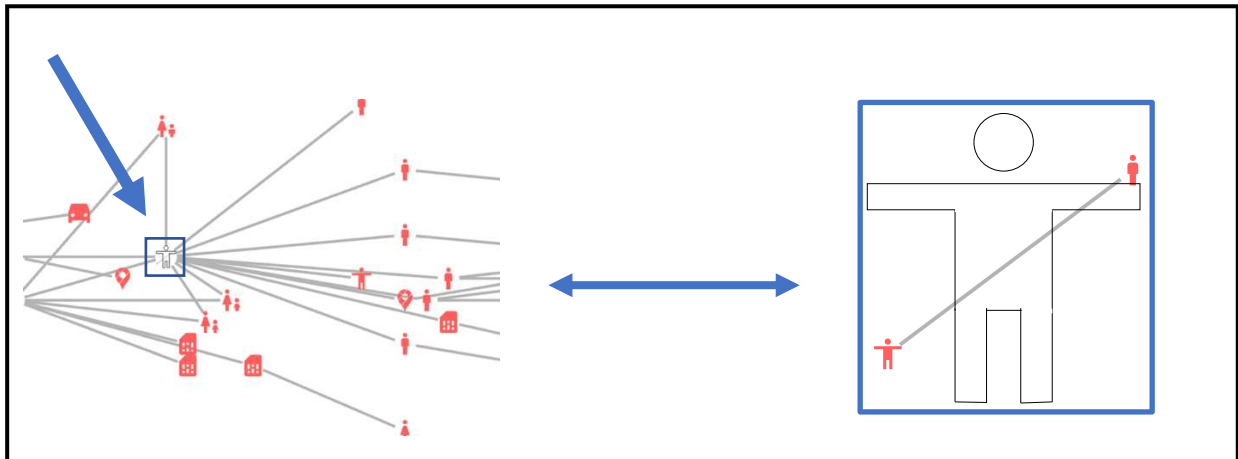


Figure 22 Illustration d'un méta-nœud (icône blanc) constitué de deux identités distinctes (icône rouges) appartenant à la même personne

Enfin, le graphe relationnel de travail est obtenu (figure 23).

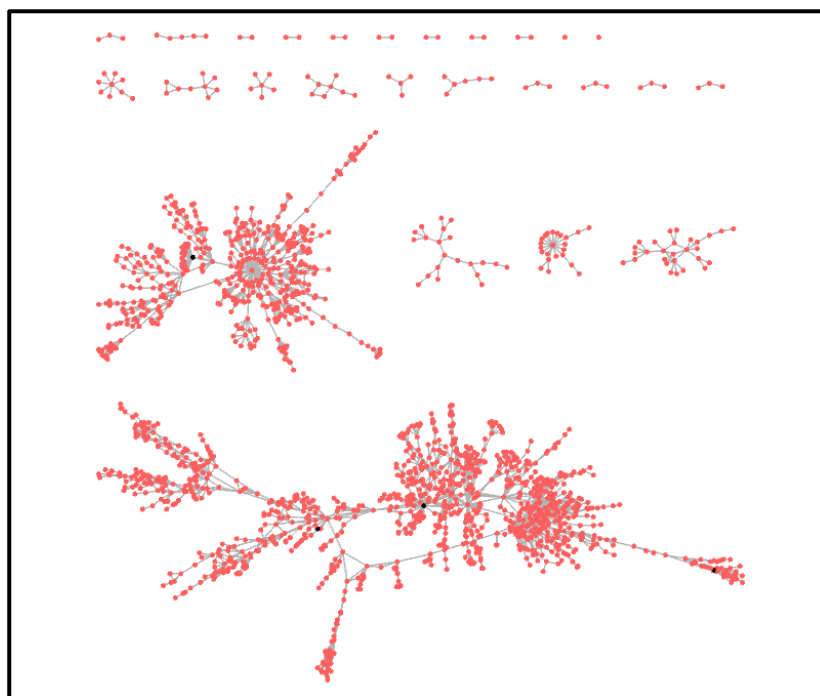


Figure 23 Graphe relationnel, redessiné avec l'algorithme FM³³²³

³²³ Algorithme de dessin de graphe optimisant les distances entre les entités pour éviter les superpositions graphiques afin d'améliorer la visualisation une fois le(s) graphe(s) généré(s).

b) Détection graphique d'anomalies dans les données

Outre la représentation et l'analyse de données, la visualisation graphique permet la détection d'anomalies non repérées lors des corrections du fichier de données. Cette étape a notamment permis de repérer que certains liens contenaient toujours « susceptible de » dans leur libellé³²⁴, ou que certaines icônes ne correspondaient pas aux libellés des entités auxquelles ils étaient associés (par exemple une icône « restaurant » associée au libellé « LIEU PRIVÉ » au lieu de « LIEU PUBLIC »³²⁵). Ainsi, le processus de nettoyage et uniformisation décrit à la section précédente a été effectué de manière itérative, c'est-à-dire qu'après avoir effectué les corrections manuelles, les données ont été importées dans *Tulip* ; la visualisation a révélé que certaines anomalies n'avaient pas été repérées lors de la première phase : les modifications adéquates ont donc été effectuées dans le fichier unique, qui a ensuite été à nouveau importé, révélant d'autres anomalies. Ce processus a été répété jusqu'à ce qu'aucune erreur ne soit plus détectée. C'est ainsi sur la onzième version du fichier unique que nous avons travaillé dans la suite de ce travail.

Les modifications effectuées suite à la visualisation graphique ont été de deux natures : elles concernaient d'une part le nettoyage et l'uniformisation du fichier unique décrit dans la section précédente : détection de nœuds multiples pour la même entité (fautes d'orthographe, de frappe), de catégories multiples pour la même entité (plusieurs icônes associées à l'entité) ... D'autre part, la visualisation a permis de mettre en évidence des arêtes multiples non pertinentes au sens des analyses proposées par la suite car elles apporteraient des biais au modèle. Les arêtes dont il s'agit concernaient des liens fixes référencés plusieurs fois dans les données, tels qu'une relation de domiciliation entre une personne donnée et un lieu donné. En effet, pour que la méthode d'analyse appliquée dans la suite soit valable, il est nécessaire que seuls les liens correspondant à des événements ponctuels se répétant soient représentés plusieurs fois (une fois par occurrence de l'événement). La notion de liens fixe ou ponctuel est basée sur la sémantique de la catégorie de lien, c'est-à-dire qu'un lien de catégorie « adresse » représente une relation inscrite dans la durée, tandis qu'un lien de catégorie « localisation » représente une relation ponctuelle et donc répétable. Lorsqu'un lien fixe existait plusieurs fois, une seule occurrence de ce dernier a été conservée dans les données. Le tableau 35 illustre cette démarche.

³²⁴ Cf. tableau 26 et 27.

³²⁵ Cf. tableau 28.

Tableau 35 Illustration de la gestion des liens fixes et ponctuels

ENTITE SOURCE	LIBELLE LIEN	DATE LIEN	ENTITE DESTINATION	EVENEMENT	LIGNE
INDIVIDU A	Adresse	04/10/2016	1 rue Villemain	Fixe	CONSERVEE
INDIVIDU B	Localisation	03/02/2016	1 rue Villemain	Ponctuel	CONSERVEE
INDIVIDU B	Localisation	07/12/2016	1 rue Villemain	Ponctuel	CONSERVEE
INDIVIDU A	Adresse	∕	1 rue Villemain	Fixe	SUPPRIMEE
INDIVIDU B	Localisation	20/07/2017	1 rue Villemain	Ponctuel	CONSERVEE

La figure 24 illustre ce processus itératif de pré-traitements et de visualisations graphiques jusqu'à l'obtention de données fin prêtes pour l'analyse.

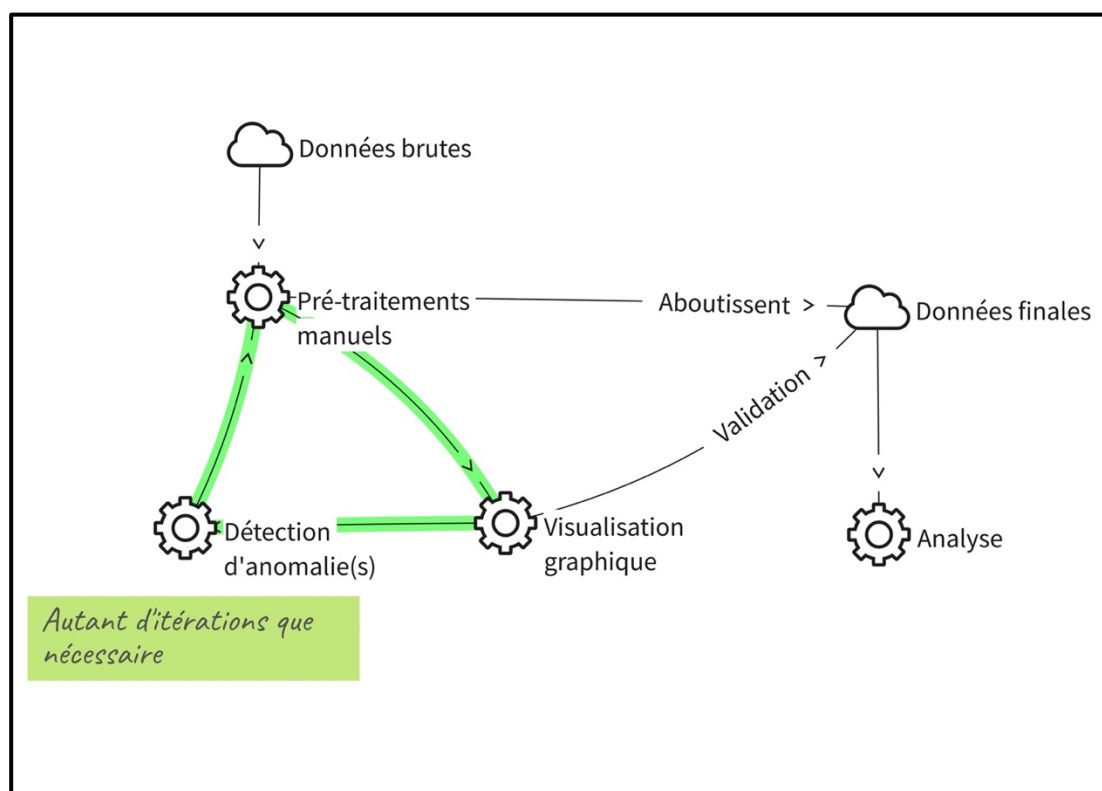


Figure 24 Processus de détection et correction des anomalies dans les données

Dans la suite de ce travail, nous nommons « graphe relationnel *n-mode* » le graphe importé dans *Tulip* après nettoyage, standardisation et homogénéisation des données. Le graphe relationnel *n-mode* possède seize composantes connexes, *i.e.* indépendantes les unes des autres, 1 743 nœuds dont 673 sommets de type « PERSONNE » pour un total de dix types d'entités, et 2 357 arêtes pour 19 types de lien.

CHAPITRE V MÉTHODOLOGIE GÉNÉRALE : DU GRAPHE MULTIVARIÉ AU RÉSEAU SOCIAL

V.1 Extraction sémantique du réseau social depuis le graphe multivarié : approche par la science des réseaux

Le premier objectif du présent travail est de proposer une analyse structurelle et fonctionnelle du réseau social constitué par l'organisation criminelle étudiée. Dans un second temps, il s'agira de développer une méthode d'analyse des réseaux criminels permettant aux enquêteurs d'identifier rapidement les membres d'une organisation essentiels à son fonctionnement – de par leurs positions et/ou fonctions clefs, afin d'entraver efficacement ses activités criminelles.

Les données, décrites au chapitre III « Le cas d'usage et les données d'étude » et dans la section IV.1 « Préambules de l'analyse : pré-traitements et traitements des données » sont fournies par les actes d'enquête effectués par les gendarmes : filatures, géolocalisations de véhicules, mises sur écoute de lignes téléphoniques, perquisitions de lieux et de véhicules... Ces données, initialement recueillies et stockées sur des procès-verbaux de procédure, ont ensuite été saisies par l'analyste dans une matrice d'analyse sous forme d'*edge list*, puis nettoyées et formatées par la doctorante.

a) Nécessité

Comme illustré dans la section précédente, les entités ainsi que les liens sont de différentes natures et peuvent être classés en catégories générales : personne, lieu, téléphone, véhicule..., pour les nœuds ; propriété, utilisation, domiciliation, communication... pour les liens.

Ainsi, elles ne représentent pas uniquement des liens directs entre des protagonistes impliqués dans l'affaire, mais aussi et surtout des liens entre les différentes catégories d'entités : le graphe relationnel final n'est donc pas un sociogramme mais un graphe multivarié, également désigné comme « graphe *n-mode* » dans la suite. Deux niveaux de zoom du graphe final sont présentés en figure 25.

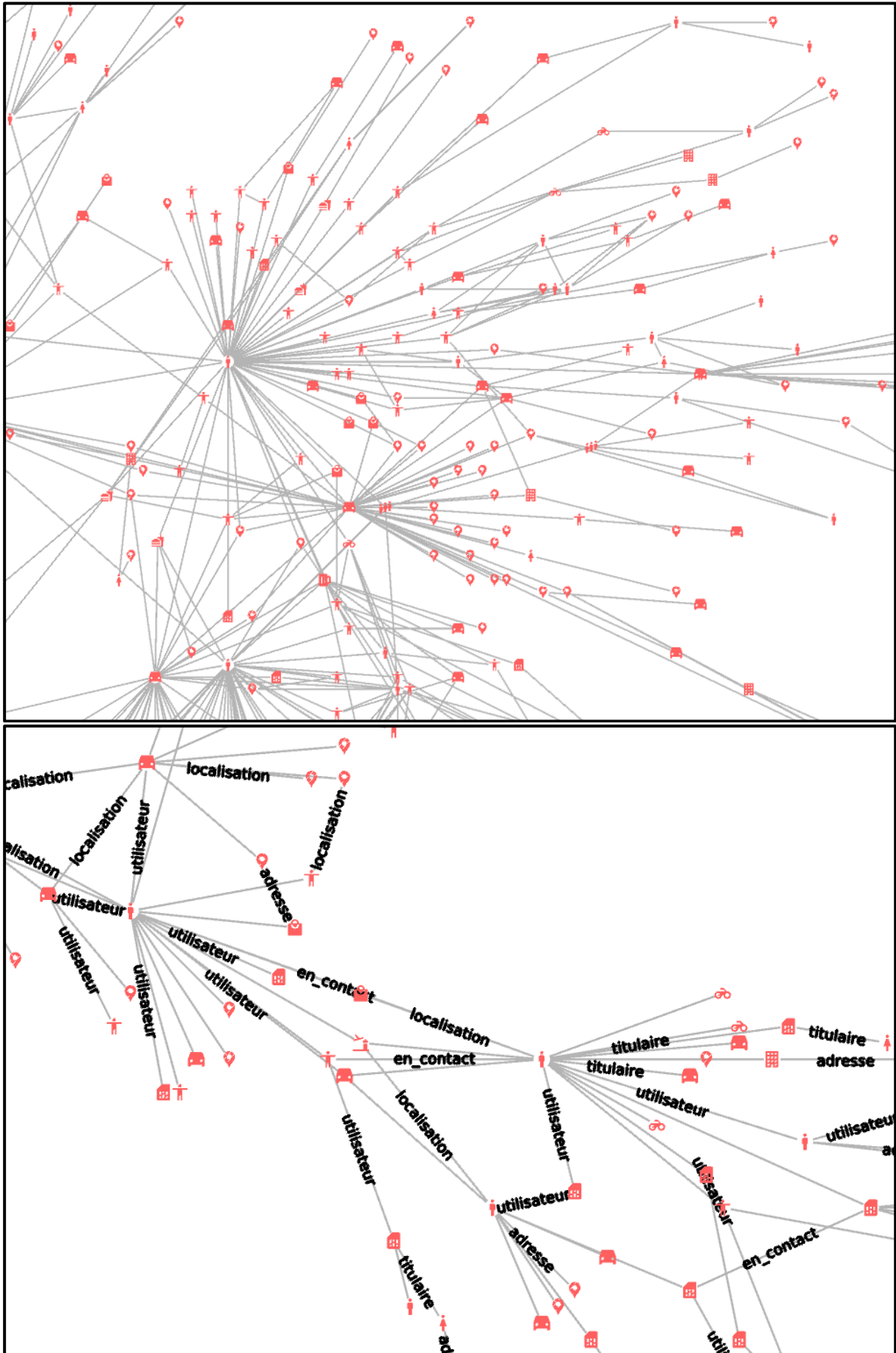


Figure 25 Extraits zoomés du graphe *n-mode* nettoyé

Tulip permet l'analyse de ce graphe du point de vue de la théorie des graphes en appliquant des calculs de distances, mesures de centralités, de densité, de connectivité, en identifiant des clusters dans ces données... Cependant, l'interprétation de telles mesures donnerait des résultats bien éloignés de la problématique d'identification d'individus clefs, dans la mesure où l'analyse permettrait en fait d'identifier tous types d'entités centrales dans le fonctionnement du réseau criminel. Or, appliqué au contexte de la recherche, de tels résultats ne seraient pas pertinents : la problématique n'est pas simplement d'apporter un soutien à l'enquête en identifiant ses éléments centraux, mais bien de permettre une détection des acteurs de l'organisation nécessaires à son fonctionnement.

Ainsi, afin d'effectuer des analyses et d'obtenir des résultats répondant à la problématique, il est nécessaire d'appliquer les mesures d'ARS non pas sur le graphe relationnel *n-mode*, mais sur le réseau social – ou sociogramme – de l'organisation criminelle. Les diverses mesures appuieront alors la mise en évidence d'individus clefs.

Dès lors, une transformation de modèle doit être opérée pour extraire du graphe *n-mode* le graphe relationnel (*one-mode*) représentant des connexions directes entre individus indirectement matérialisées dans les données.

Le graphe *one-mode* des individus impliqués dans l'affaire ainsi obtenu constituera par la suite le matériel de travail sur lequel seront effectuées les analyses structurelles et fonctionnelles du « champ criminel » que constitue l'organisation.

b) Méthodologie de transformation : inspiration issue des concepts de grammaires formelles

Le cas présent est non trivial, car la structure des données du cas d'usage diffère des cas usuels de transformation de modèle.

En effet, les méthodes classiques de transformation de modèles permettent l'extraction d'un graphe *one-mode* depuis un graphe *two-mode*, où seulement deux types d'entités sont représentés (par exemple des individus et des événements auxquels ces individus ont assisté). En outre, les liens présents dans les cas classiques représentent des connexions inter-entités uniquement, c'est-à-dire qu'il n'existe pas de liens reliant directement deux entités de même type (individu-individu ou événement-événement), contrairement au présent cas d'usage³²⁶ : lors d'une enquête, les gendarmes recueillent toutes les données relationnelles que les

³²⁶ Cf. section II.4 a) « Transformations de modèles : application de la théorie des graphes ».

différentes techniques d'enquête fournissent, indifféremment de leur structure, y compris des connexions intra-entités, comme deux téléphones en communication ou deux véhicules repérés côte à côte.

Pour ces deux raisons, le cas d'usage diffère fondamentalement des cas classiques, rendant nécessaire le développement d'une nouvelle méthode de transformation de modèle.

Les différentes étapes de la méthodologie appliquée ont ainsi été élaborées et réalisées progressivement, au cours de multiples sessions de travail avec Bruno Pinaud.

La question qui se pose est la suivante : parmi l'ensemble des chemins reliant graphiquement deux entités de type « PERSONNE » dans le graphe *n-mode*, lesquels permettent d'inférer sur l'existence sur le terrain d'une connexion directe entre les personnes sources et destinations de ces chemins ? En théorie des graphes, un chemin est une chaîne de nœuds et d'arêtes distinct(e)s les un(e)s des autres permettant d'aller d'un point du graphe à un autre sans discontinuité³²⁷. Nous nous intéressons ici uniquement aux chemins dont les nœuds source et destination sont de type « PERSONNE ». Dans la suite de ce travail, les nœuds et les arêtes seront définis par leur catégorie sémantique (non plus avec leurs libellés initiaux³²⁸). Seront donc nommées « chemin » les chaînes de catégories de nœuds et de catégories d'arêtes permettant d'aller d'un nœud de catégorie « PERSONNE » à un autre. Ce changement d'échelle est illustré par les figures 26 et 27.

³²⁷ Stanley Wasserman et Katherine Faust, *Social Network Analysis Methods and Applications*, *op. cit.*

³²⁸ Cf. tableaux 28 à 31, pages 107 à 109.

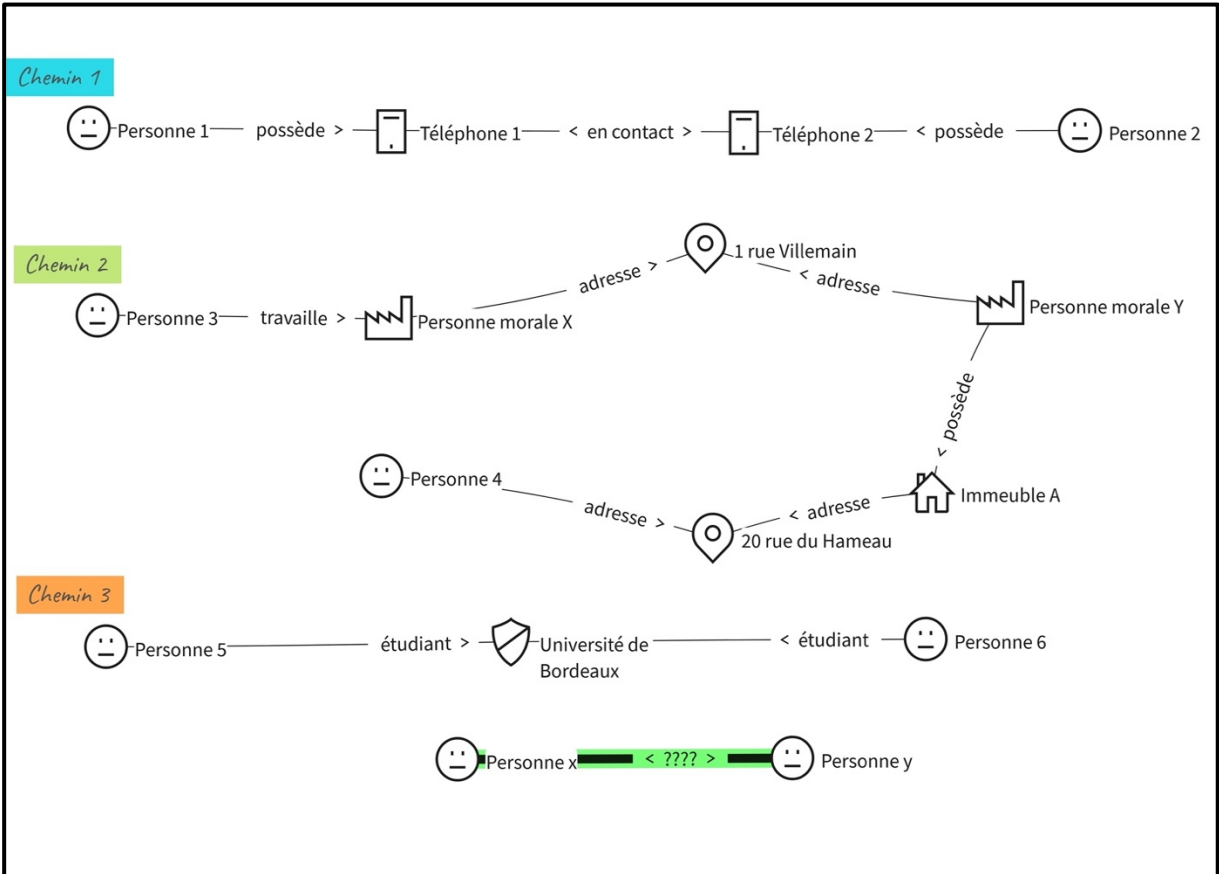


Figure 26 Exemples de chemins reliant graphiquement deux personnes (libellés initiaux)

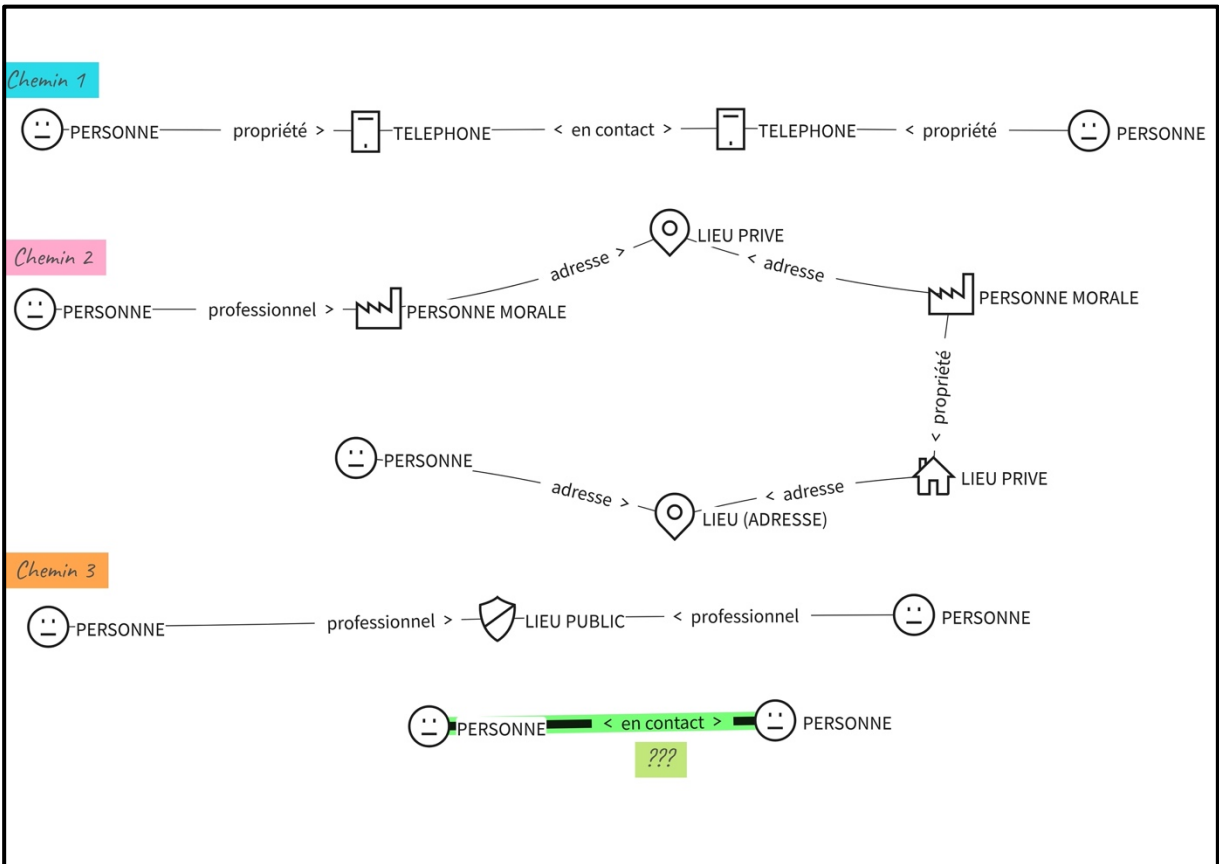


Figure 27 Changement d'échelle sémantique pour utiliser les catégories de liens et d'entités (passage des libellés aux catégories)

Nous proposons d'étudier cette question en s'inspirant de méthodes issues du Traitement Automatique des Langues Naturelles³²⁹ (TALN) combinées à l'approche par la science des réseaux. Le TALN est une discipline à la frontière entre la linguistique et l'intelligence artificielle, rassemblant l'ensemble des méthodes et outils donnant la capacité à un programme informatique de comprendre le langage tel qu'il est construit. C'est au cours d'une discussion passionnée avec Nicolas Merli, ingénieur linguiste, qu'est née l'idée d'appliquer des concepts de grammaires formelles.

Une grammaire est un formalisme permettant de définir une syntaxe et ainsi un langage formel, tel qu'un ensemble de mots admissibles sur un alphabet donné. Le vocabulaire en grammaire formelle – appliqué à notre cas (figure 36) – peut être décomposé et défini comme suit :

1. Un ensemble fini de symboles terminaux, qui sont les plus petits éléments du langage formalisé et donc irréductibles.

Dans notre cas, il s'agit de l'ensemble des trinômes {CATEGORIE ENTITE - catégorie lien – CATEGORIE ENTITE} existant dans les données.

2. Un ensemble fini de symboles non terminaux, résultant de la combinaison de terminaux et non terminaux selon les règles de réécriture qui régissent le langage.

Dans notre cas, il s'agit de l'ensemble des chemins reliant deux nœuds de type « PERSONNE ».

3. Un ensemble de règles de réécriture régissant les modalités d'association des terminaux et non terminaux.

Dans notre cas, ces règles sont de deux natures : une fonction de fusion des terminaux correspondant à des arêtes multiples et une fonction d'agrégation des terminaux sur l'ensemble des chemins.

Tableau 36 Illustration du vocabulaire emprunté aux grammaires formelles

Vocabulaire	Symbole terminal (trinôme)
Forme	{CATEGORIE ENTITE - catégorie lien – CATEGORIE ENTITE}
Exemple	{PERSONNE – propriétaire – VÉHICULE}.
Vocabulaire	Symbole non terminal (combinaison de trinômes)
Forme	{PERSONNE – cat lien1 – CAT Ei1 – cat lien2 – CAT Ei2 – cat lien3 – PERSONNE} ³³⁰
Exemple	{PERSONNE – propriété – VHL – localisation – LIEU PRIVÉ – adresse – PERSONNE}

³²⁹ Aussi connu dans la communauté sous le nom anglais *Natural Language Processing (NLP)*.

³³⁰ Pour des raisons d'espace et de lisibilité, les abréviations suivantes sont ici utilisées :

cat = catégorie ;

Ei = ENTITÉ Intermédiaire.

L'idée est de condenser les symboles non-terminaux (les chemins, qui sont des combinaisons de symboles terminaux) selon les règles de réécriture pour obtenir le symbole terminal qui nous intéresse, à savoir {PERSONNE – en contact – PERSONNE}. L'objectif est de calculer, sur la base de la sémantique des entités et liens intermédiaires, le niveau de confiance que chaque chemin confère en l'existence d'une connexion directe entre les personnes sources et destination, *i.e.* en la validité de l'hypothèse selon laquelle le symbole terminal {PERSONNE – en contact – PERSONNE} peut être considéré dans la suite du travail.

c) Définition de la notion de « force liante » d'un symbole

La méthode proposée, qui s'inspire des concepts de grammaires formelles, consiste à attribuer à chaque symbole terminal {CATEGORIE ENTITE - catégorie lien – CATEGORIE ENTITE} existant dans les données une pondération sémantique, c'est-à-dire une valeur chiffrée comprise entre 0 et 1 reflétant ce que nous appellerons la « force liante » du trinôme.

La « force liante » d'un symbole terminal est définie comme la capacité du trinôme, en tant que fragment d'un chemin liant deux personnes, à soutenir l'existence du lien direct entre ces personnes induit par le chemin.

La « force liante » d'un symbole non terminal (*i.e.* d'un chemin) est définie comme sa capacité à soutenir l'existence du lien direct qu'il induit. La force liante d'un symbole non terminal est calculée en combinant, selon les règles de réécriture, les forces liantes de l'ensemble des symboles terminaux qui le composent, et reflète le niveau de confiance qui peut être accordé à l'existence du symbole {PERSONNE – en contact – PERSONNE} ainsi obtenu.

Pour bien comprendre ces notions, et la définition de notre concept de « force liante », nous proposons l'analogie suivante (figure 28).

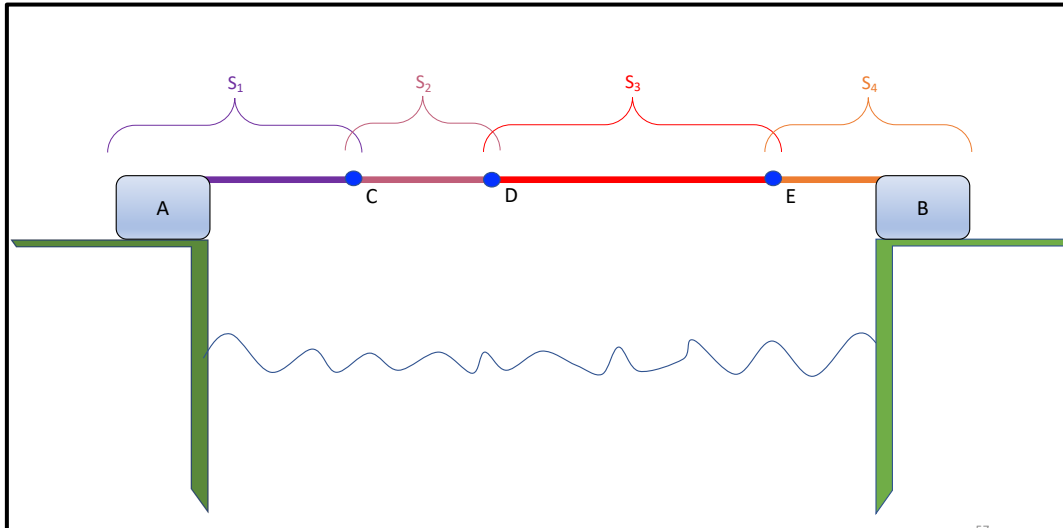


Figure 28 Analogie explicative de la notion de force liante

Un pont est construit au-dessus d'une rivière, qui relie les rives A et B. Ce pont est composé de 4 fragments (S_1 , S_2 , S_3 et S_4) juxtaposés. Chaque fragment est composé d'une attache (A, B, C, D, E) et d'un tronçon (violet, mauve, rouge, orange). Connaissant la solidité de chaque fragment, on cherche à déterminer la solidité du pont, c'est-à-dire si le pont permet d'aller de la rive A à la rive B sans risque. La solidité globale du pont dépend alors de la solidité individuelle de chacun des fragments qui le composent, mais aussi de la manière dont ces fragments sont agencés entre eux et du nombre de fragments total.

Dans notre cas d'usage, les rives A et B sont des personnes, les attaches C, D et E correspondent aux entités intermédiaires, les tronçons aux arêtes, les fragments aux trinômes, et le pont au chemin graphique reliant A et B. La solidité des fragments correspond à la force liante des symboles terminaux, et la solidité du pont à la force liante des symboles non terminaux.

Le modèle développé cherche à se rapprocher des raisonnements cognitifs humains sur la base desquels l'analyste décide s'il peut établir la connexion directe ou non.

Pour reprendre l'exemple de la figure 26, l'analyste, en lisant les données, va les interpréter sémantiquement, et décider de considérer que le chemin 1 permet de conclure à une connexion directe entre les personnes 1 et 2, et que les chemins 2 et 3 ne fournissent pas d'éléments assez solides pour établir une connexion entre les personnes 3 et 4 ni 5 et 6. Il ne s'agit pas d'affirmer que le lien direct n'existe pas, mais de poser que les données à disposition ne permettent pas de considérer que la connexion existe. Ainsi, les forces liantes des symboles non terminaux que sont les chemins 1, 2 et 3 illustreront le niveau de confiance que l'analyste peut avoir dans l'existence d'une connexion entre les deux individus concernés par chacun des chemins :

$$P(\text{chemin 1}) > P(\text{chemin 3}) > P(\text{chemin 2})$$

Enfin, la prise de décision, *i.e.* l'acceptation ou le rejet du lien direct, sera effectuée sur la base d'une valeur seuil S qui sera définie au chapitre VI.

Les sections suivantes détaillent la méthode d'extraction du réseau social de l'organisation et son application sur le cas d'usage.

d) Pré-traitements graphiques : réduction de l'espace de recherche

Ainsi, en premier lieu, une extraction de l'ensemble des chemins graphiques permettant de relier entre elles chaque paire d'entités de type « PERSONNE » du graphe a été envisagée afin de les analyser selon les principes de grammaires formelles, *i.e.* en découpant les chemins en symboles terminaux puis en les reconstruisant selon les règles de réécriture définies.

L'intérêt de la méthode est de rechercher l'ensemble des chemins reliant une paire de personnes données plutôt que de rechercher le plus court chemin³³¹ comme c'est classiquement le cas en analyse des réseaux sociaux (ARS), notamment à l'aide de l'algorithme de Dijkstra³³². En effet, les chemins doivent être traités sur la base de la sémantique des chaînes de liens qui les décomposent avant d'être traités par une approche mathématique. Classiquement, les calculs de distance en ARS sont effectués sur des réseaux sociaux, *i.e.* des graphes illustrant des relations entre personnes. Ici, l'objectif est de construire un réseau social sur la base de données multiples, où tous les liens ne sont pas équivalents d'un point de vue sémantique : ils doivent donc être tous considérés.

³³¹ La longueur d'un chemin entre deux nœuds correspond au nombre d'arêtes intermédiaires par lesquelles le chemin passe pour aller du nœud source au nœud destination.

³³² Stephen P. Borgatti, Martin G. Everett et Jeffrey C. Johnson, *Analyzing social networks*, *op. cit.*

En théorie des graphes, l'algorithme de Dijkstra est utilisé pour résoudre le problème dit « du plus court chemin ». Cet algorithme calcule des plus courts chemins à partir d'une source vers tous les autres nœuds d'un graphe, orienté ou non. Il peut aussi être utilisé pour calculer un plus court chemin entre un nœud de départ et un nœud d'arrivée.

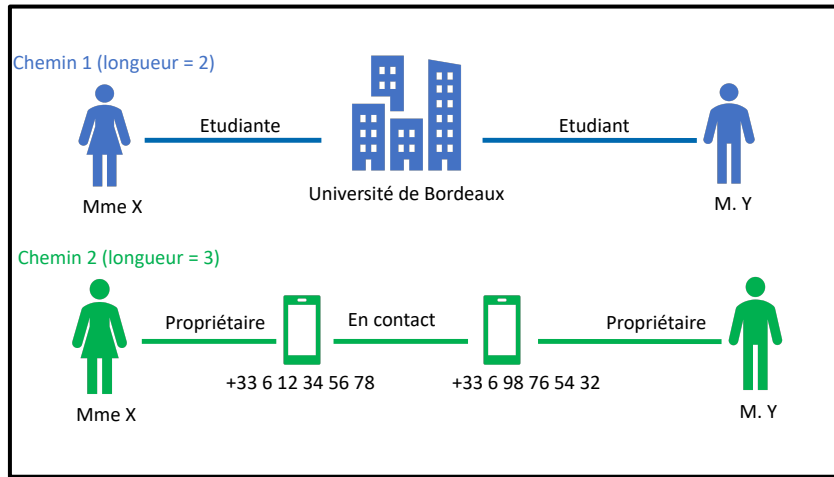


Figure 29 Illustration de l'indépendance de la sémantique et de la longueur des chemins.

Ainsi, pour une même paire d'individus, un chemin long peut avoir une force liante supérieure à celle d'un chemin court, comme l'illustre la figure 29. Dans cet exemple, le chemin 2 est sémantiquement plus significatif que le chemin 1, c'est-à-dire que sa sémantique confère un niveau de confiance dans l'existence du lien direct entre Mme X et M. Y supérieur à celui du chemin 1, bien que ce dernier soit plus court. En outre, si plusieurs chemins graphiques résultant en un lien direct avec un niveau de confiance acceptable existent entre deux individus donnés, il est important de prendre en compte cette multiplicité, raison pour laquelle, pour chaque paire de personnes, l'ensemble des chemins graphiques existants doivent être extraits.

Nous avons cependant fixé comme condition initiale le fait que les entités de type « PERSONNE » ne peuvent être considérées par l'algorithme que comme entité sources et destinations des chemins à extraire, mais jamais comme entités intermédiaires. Ainsi, si lors du parcours d'un chemin graphique entre deux personnes l'algorithme rencontre un nœud de type « PERSONNE », il terminera le chemin sur ce dernier en le fixant comme destination plutôt que de continuer le calcul en passant au nœud suivant, comme schématisé dans la figure 30.

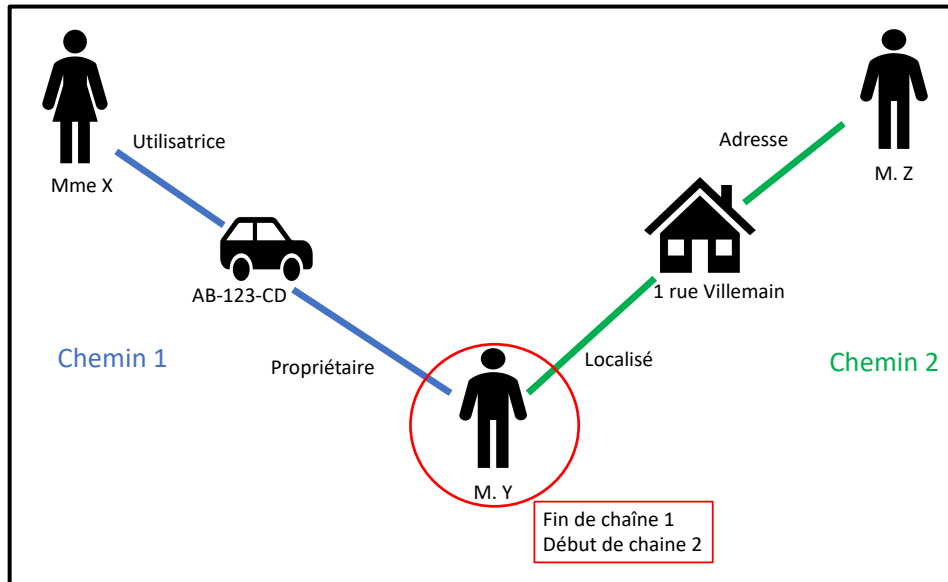


Figure 30 Les nœuds de type « PERSONNE » ne peuvent être des nœuds intermédiaires

Cette décision résulte directement du cas d'application des présentes analyses : nous ne cherchons pas simplement à extraire un réseau social « classique », mais bien le réseau social d'une organisation criminelle. Dans un réseau social « classique », Granovetter définit la « triade interdite » (figure 31) comme une configuration impossible : si A possède un lien fort avec B et avec C, alors il n'est pas possible que B et C n'aient pas de lien³³³.

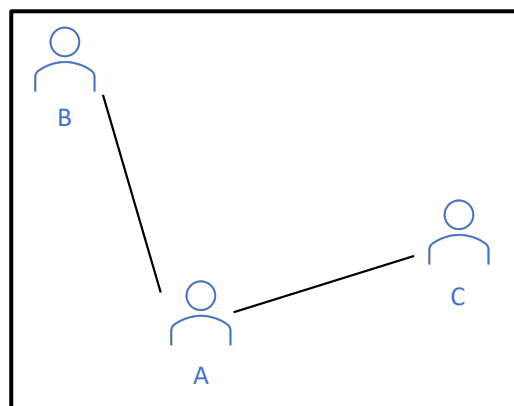


Figure 31 Triade interdite de Granovetter

Or, dans un réseau criminel, les protagonistes, afin de conserver à la fois leur importance stratégique et leurs éventuelles commissions, n'ont aucun avantage à faire bénéficier les autres membres de l'organisation de leurs contacts – ils ont tout intérêt à rester un intermédiaire essentiel. De plus, la compartimentation est l'un des six principes de survie des organisations clandestines³³⁴ : en réduisant le nombre d'individus ayant accès aux autres membres, plans ou opérations, ce principe aide à protéger le réseau dans son ensemble.

³³³ Mark S. Granovetter, « The Strength of Weak Ties », *American Journal of Sociology*, vol. 78, n° 6, 05/1973, pp. 1360-1380.

³³⁴ United States DoD, *Countering Threat Networks*, CreateSpace Independent Publishing Platform, 2016, 164 p.

Ainsi, a été fait le choix méthodologique de considérer que, dans l'exemple de la figure 31, le fait que B soit lié avec A, lui-même lié avec C ne permet pas de conclure sur l'existence d'une relation directe entre B et C – et ce, quelle que soit la force sémantique des chaînes de liens.

Compte tenu du volume de données à disposition (1 743 nœuds, dont 673 personnes, pour 2 357 arêtes), l'extraction de l'intégralité des chemins reliant graphiquement des paires de d'individus³³⁵ représente plusieurs centaines de millions de chemins à parcourir, et plusieurs dizaines de millions de chemins à extraire. Une telle extraction n'est techniquement pas faisable, même en s'appuyant sur les capacités des ordinateurs et serveurs de l'Université de Bordeaux auxquels a accès le LaBRI. Il a donc été nécessaire d'effectuer en préambule plusieurs étapes de réduction de « l'espace de recherche »³³⁶, c'est-à-dire de diminution du nombre de possibilités que l'algorithme doit explorer pour l'extraction.

La hiérarchie de graphe de *Tulip* a été utilisée tout au long des traitements en ne travaillant que sur des sous-copies du graphe parent : à chaque étape ont été créés un ou plusieurs sous-graphes sur lesquels les algorithmes ont été appliqués sans altérer le graphe parent. Le graphe *n-mode* obtenu à l'issue des traitements décrits dans la section IV.1 (intitulé dans la suite *graph_root*) constitue la racine de la hiérarchie. La figure 32 illustre la hiérarchie des sous-graphes générés ; les étapes en sont détaillées dans la suite.

³³⁵ À l'exception de ceux contenant des intermédiaires de type « PERSONNE ».

³³⁶ Mauricio Abel Gomez Soto, « Quelques propriétés topologiques des graphes et applications à Internet et aux réseaux », *Réseaux et télécommunications*, Paris Diderot University, 2011, p. 130.

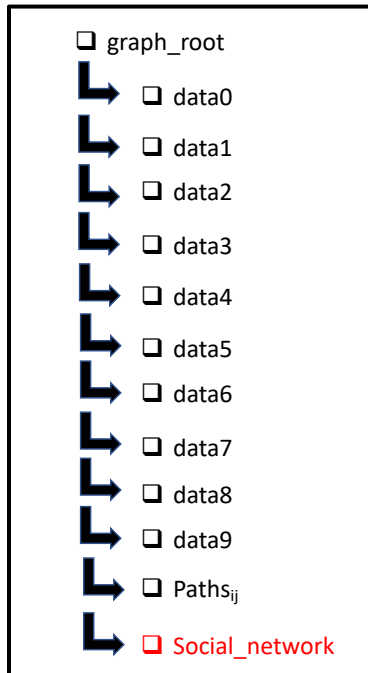


Figure 32 Schéma de la structure en sous-graphes utilisée avec *Tulip* pour réduire la complexité du graphe

Le premier sous-graphe, *data0*, est une copie de la racine, *graph_root*, dans laquelle n'ont pas été importées les arêtes des trinômes {nœud – arête – nœud} dont la force liante est nulle³³⁷. Il contient 1 743 nœuds pour 2 345 arêtes. À partir de *data0*, les arêtes multiples entre deux entités ont été fusionnées (figure 33) en une arête unique dans le sous-graphe *data1*³³⁸, qui contient toujours 1 743 nœuds, mais pour 2 100 arêtes.

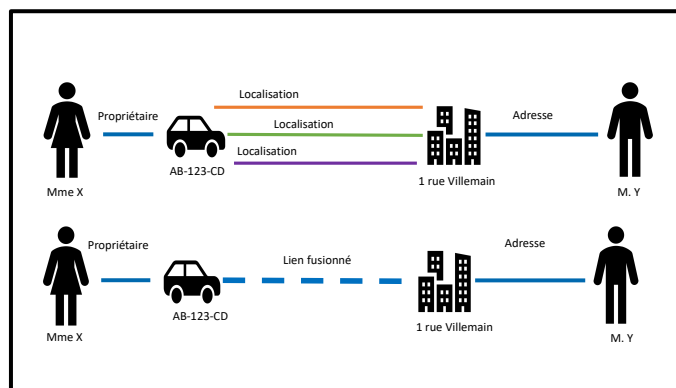


Figure 33 Principe de fusion de liens multiples

Le graphe *data2* est une copie de *data1* sur laquelle ont été appliqués les traitements suivants : suppression des composantes connexes³³⁹ du graphe ne contenant pas au moins deux nœuds de type « PERSONNE » et retrait des entités de bout de chaîne non pertinentes, *i.e.* tous

³³⁷ Voir section V.2 a) « Attribution des valeurs de force liante des symboles terminaux. »

³³⁸ Les règles de fusion des arêtes multiples sont détaillées dans la section V.2 b) « Règles de réécriture en série et en dérivation. »

³³⁹ Est appelée « composante connexe » toute partie isolée d'un graphe, c'est-à-dire ne possédant pas de lien avec au moins l'une des autres composantes du graphe.

les nœuds de degré 1³⁴⁰ de type différent de « PERSONNE » ainsi que leur arête associée (figure 34. En effet, ces nœuds augmentent le temps de calcul tandis qu'ils n'appartiennent à aucun chemin d'intérêt, dans la mesure où ils ne sont connectés qu'à un seul nœud du graphe (ils ne peuvent ainsi être ni intermédiaires, ni source, ni destination de chemins).

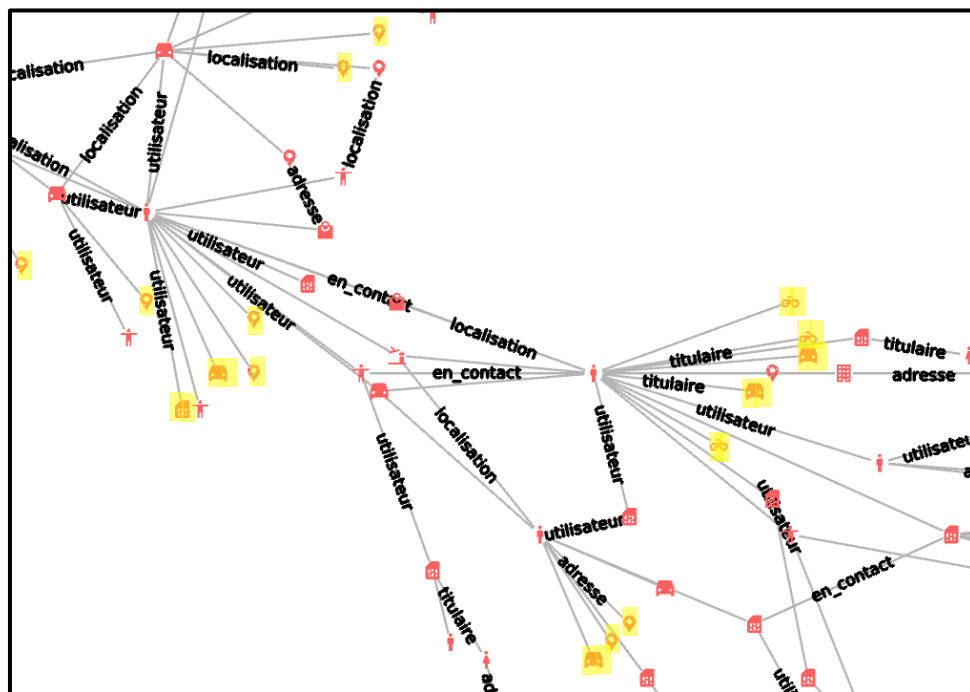


Figure 34 Retrait des nœuds et arêtes non pertinents (en jaune)

Cette étape a été répétée jusqu'à ce que tous les nœuds de degré 1 soient de type « PERSONNE » uniquement. Le graphe *data2* contient 1099 nœuds pour 1470 arêtes, soit une réduction respectivement de 36,9% et 37,3% par rapport à *data0*.

Ces premières étapes ont permis de réduire le nombre de chemins à parcourir, mais n'ont agi ni sur le nombre ni sur la complexité des chemins « personne-personne » à extraire. Malgré ces réductions, les possibilités et chemins restent indénombrables : une méthode de réduction de la complexité du graphe doit être mise en place.

e) Traitements du graphe : réduction de la complexité

Afin de réduire la complexité toujours trop importante du dernier sous-graphe obtenu, *data2*, nous avons appliqué un ensemble d'opérations visant à se rapprocher le plus possible

³⁴⁰ Le degré d'un nœud donné correspond au nombre d'arêtes qui lui sont incidentes.

d'un graphe en forme d'arbre, dans lequel il n'existe par définition qu'un seul chemin entre deux nœuds donnés³⁴¹.

Pour ce faire, nous entendons par « réduction de la complexité » deux types de manipulations : celles visant à réduire le nombre de chemins existants entre chaque paire de nœuds³⁴², et celles visant à en réduire la longueur.

Afin de réduire le nombre de chemins, une méthode de fusion des liens multiples (aussi désignés dans la suite comme « liens en dérivation ») a été mise en place. Cette étape est nécessaire dans le but d'obtenir un unique lien direct entre chaque paire de nœuds de type « PERSONNE » dont la pondération reflète l'ensemble des chemins existants dans les données. Ce processus est schématisé dans les figures 35 et 36.

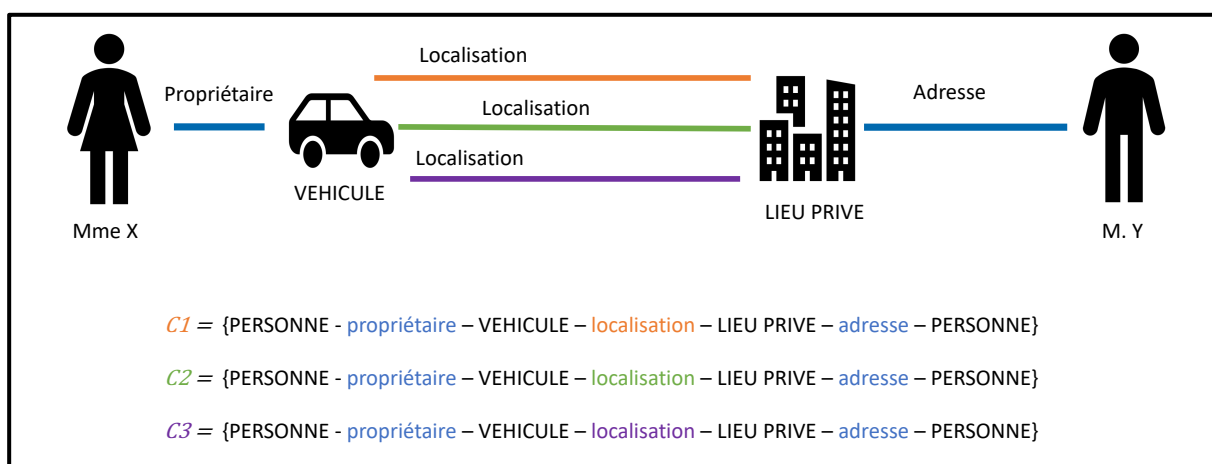


Figure 35 Fusion des liens multiples : situation initiale

Dans cet exemple, trois chemins distincts ($C1$, $C2$ et $C3$) relient Mme X et M. Y. Une méthodologie pourrait consister à calculer les forces liantes de ces trois chemins puis à les fusionner pour obtenir la pondération unique de {Mme X – en contact ? – M. Y}. Cependant, cette méthode biaiserait les calculs, car elle implique de considérer les symboles terminaux {PERSONNE – propriétaire – VÉHICULE} et {LIEU PRIVÉ – adresse – PERSONNE} dans le calcul de chacun des trois chemins. Or, ces deux trinômes, bien que communs à tous les chemins, n'ont qu'une seule occurrence dans les données pour la paire de personnes {Mme X – M. Y}. Ainsi, il est impératif de réunir les trois arêtes en dérivation entre les nœuds « véhicule » et « lieu privé » en un lien fusionné unique pour obtenir un seul chemin C avant de pouvoir calculer sa force liante. L'opération de fusion a été effectuée selon la règle de réécriture

³⁴¹ Stanley Wasserman et Katherine Faust, *Social Network Analysis Methods and Applications*, op. cit.

³⁴² Il ne s'agit pas ici de réduire l'espace de recherche en agissant sur le nombre de chemins que doit parcourir l'algorithme (comme dans la section précédente), mais bien de réduire le nombre de chemins que l'algorithme doit extraire.

de fusion préalablement définie et détaillée dans la section CHAPITRE V b) « Règles de réécriture en série et en dérivation ».

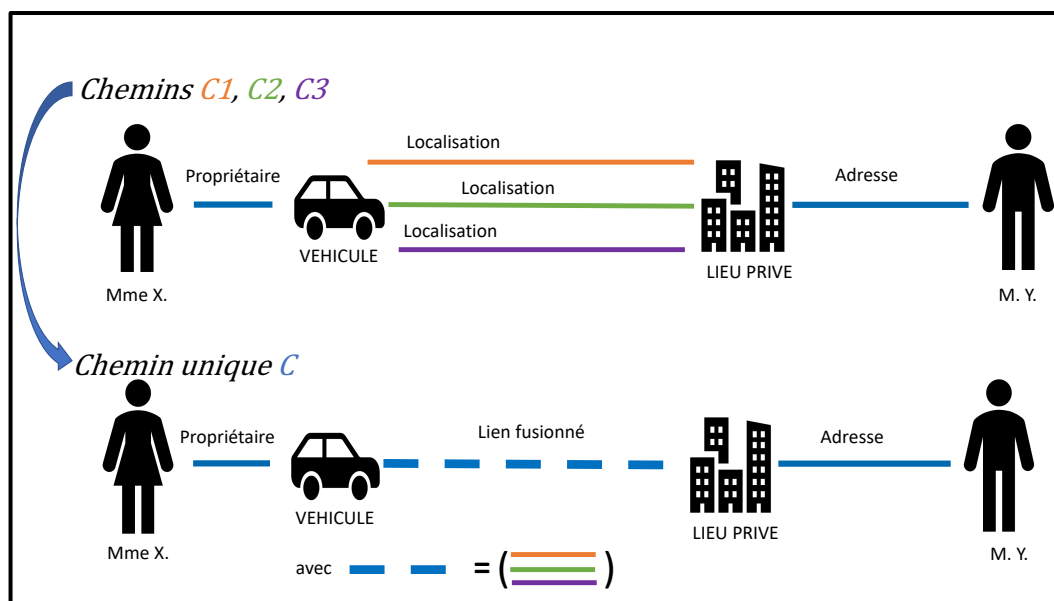


Figure 36 Schéma de la méthode de fusion des arêtes en dérivation

La deuxième opération, désignée dans la suite comme « agrégation des arêtes en série » consiste à réduire la longueur des chemins en agrégeant entre eux les trinômes {CATEGORIE ENTITE – catégorie lien – CATEGORIE ENTITE} successifs³⁴³ qui composent un chemin complet.

La méthode consiste à condenser les chemins en générant de nouvelles arêtes, appelées « méta-arêtes », dans lesquelles sont stockées les informations contenues dans les trinômes intermédiaires, comme illustré dans la figure 37.

³⁴³ Par opposition aux trinômes multiples qui sont fusionnés, les trinômes successifs sont agrégés entre eux.

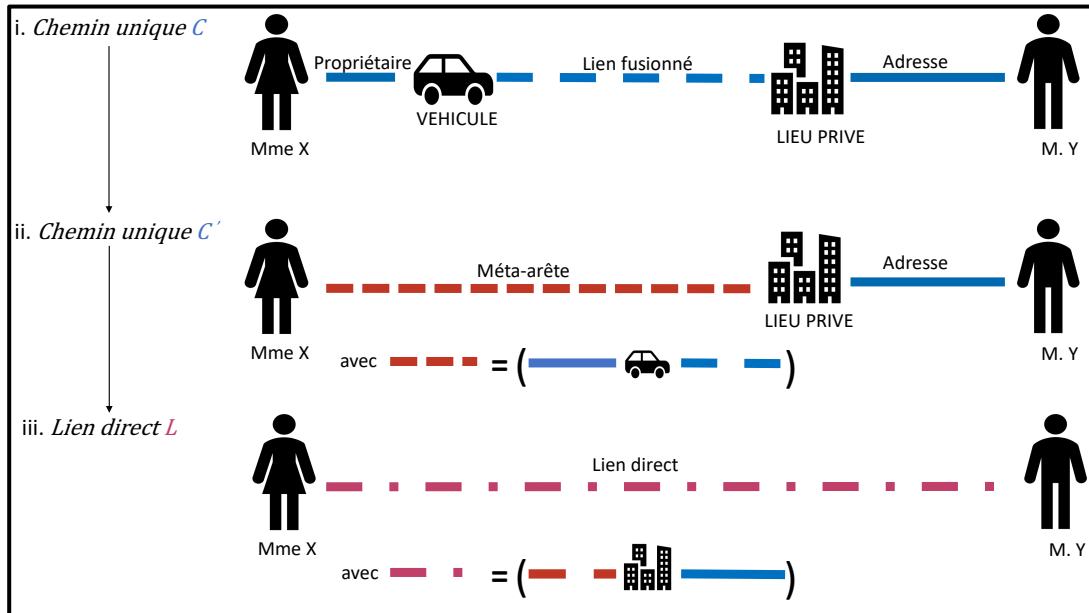


Figure 37 Schéma de la méthode d'agrégation des arêtes en série en méta-arêtes

Cette étape d'agrégation n'est appliquée qu'aux nœuds de degré 2^{344} qui ne sont pas de type « PERSONNE ». En effet, il ne serait pas pertinent de supprimer les nœuds « PERSONNE », qui nous intéressent ici en les stockant dans des méta-arêtes. La condition de degré permet de respecter la structure initiale des données au cours des traitements. En effet, dans la situation représentée dans la figure 38, où le véhicule est de degré trois, il n'est pas possible de stocker les informations relationnelles du nœud et de ses arêtes incidentes dans une seule méta-arête. Selon la méthode appliquée, il faudrait donc générer trois méta-arêtes :

- Mme X – **Méta-arête 1** – PERSONNE MORALE avec Méta-arête 1 = {utilisation-VÉHICULE-propiétaire}
- Mme X – **Méta-arête 2** – M. Y avec Méta-arête 2 = {utilisation-VÉHICULE-utilisation}
- M. Y – **Méta-arête 3** – PERSONNE MORALE avec Méta-arête 3 = {utilisation-VÉHICULE-propiétaire}

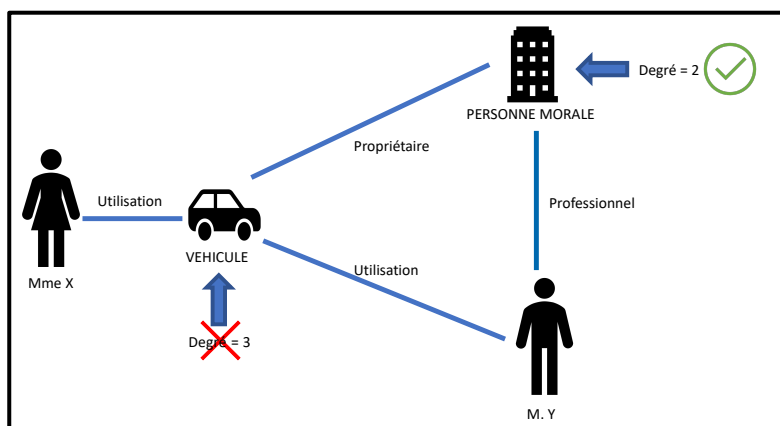


Figure 38 Conditions d'applications de l'agrégation en série (1)

³⁴⁴ Un nœud de degré 2 est un nœud connecté à deux autres entités du graphe.

Or, cela induirait un biais méthodologique non négligeable, dans la mesure où le calcul des forces liantes de chacune des deux méta-arêtes incidentes à Mme X (Méta-arêtes 1 et 2) prendrait en compte le trinôme {Mme X – utilisation – VÉHICULE} qui n'a pourtant qu'une seule occurrence dans les données (figure 39 et 40). Ainsi, dans le cas de la figure 38, l'agrégation ne pourra être effectuée qu'au niveau du nœud « personne morale », qui est bien de degré 2.

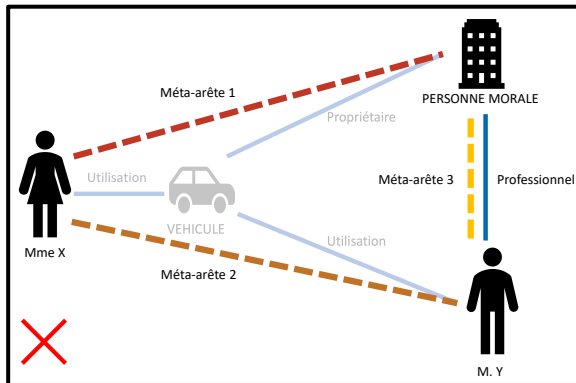


Figure 39 Conditions d'application de l'agrégation en série (2)

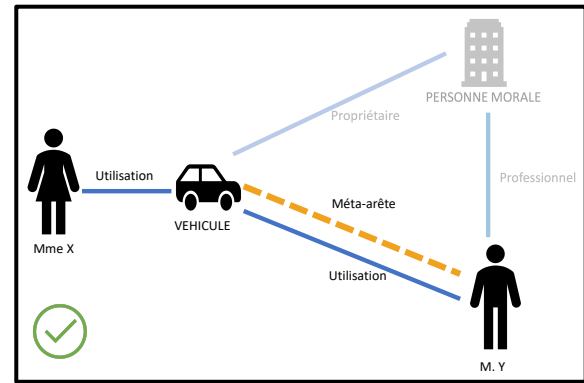


Figure 40 Conditions d'application de l'agrégation en série (3)

La démarche de réduction de la complexité ne résulte pas en une perte d'information : chaque traitement appliqué permet de réduire le nombre ou la longueur des chemins à traiter tout en conservant l'information initiale. En effet, les liens fusionnés et les méta-arêtes sont stockés sous la forme de vecteurs contenant les informations (catégories de lien et d'entités source et destination, identifiant du lien et force liante du symbole terminal) relatives aux liens en dérivation et en série dont ils sont issus. Ainsi, les liens directs uniques obtenus en fin de traitement contiennent systématiquement l'ensemble des informations des chemins initiaux correspondants. Les chemins *C*, *C'* et *L* de la figure 37 stockeront ainsi les informations représentées figures 41 à 43.

C: {PERSONNE – propriétaire – VÉHICULE – **lien fusionné** – LIEU PRIVÉ – adresse – PERSONNE}

C: {PERSONNE – propriétaire – VÉHICULE – {**localisation, localisation, localisation**} – LIEU PRIVÉ – adresse – PERSONNE} ;

Figure 41 Informations contenues dans le chemin *C*

C' : {PERSONNE - **lien agrégé** – LIEU PRIVÉ – adresse – PERSONNE}

C' : {PERSONNE – {**propriétaire – VÉHICULE – lien fusionné**}– LIEU PRIVÉ – adresse – PERSONNE}

C' : {PERSONNE – {**propriétaire – VÉHICULE – {localisation, localisation, localisation}**}} – LIEU PRIVÉ – adresse – PERSONNE} ;

Figure 42 Informations contenues dans le chemin C'

L : {PERSONNE – **lien direct** – PERSONNE}

L : {PERSONNE - {**lien agrégé – LIEU PRIVÉ – adresse**} – PERSONNE}

L : {PERSONNE – {{**propriétaire – VÉHICULE – lien fusionné**} – LIEU PRIVÉ – adresse} – PERSONNE}

L : {PERSONNE – {{**propriétaire – VÉHICULE – {localisation, localisation, localisation}**}} – LIEU PRIVÉ – adresse} – PERSONNE}.

Figure 43 Informations contenues dans le lien L

Enfin, ces opérations de fusion et d'agrégation ne sont pas effectuées successivement, mais fonctionnent de manière itérative et complémentaire, c'est-à-dire qu'elles s'alimentent mutuellement : lorsqu'une étape est terminée, l'autre débute en s'appliquant sur le sous-graphe de sortie, générant ainsi un nouveau sous-graphe de *graph_root* sur lequel la première étape opérera à nouveau, ce qui donnera lieu à une deuxième itération de la seconde étape, etc. Ces itérations se reproduisent jusqu'à ce que les deux opérations cessent de s'alimenter, c'est-à-dire lorsque le nombre de nœuds et d'arêtes cesse de diminuer. Ce processus est schématisé dans la figure 44, suivie du détail des traitements appliqués.

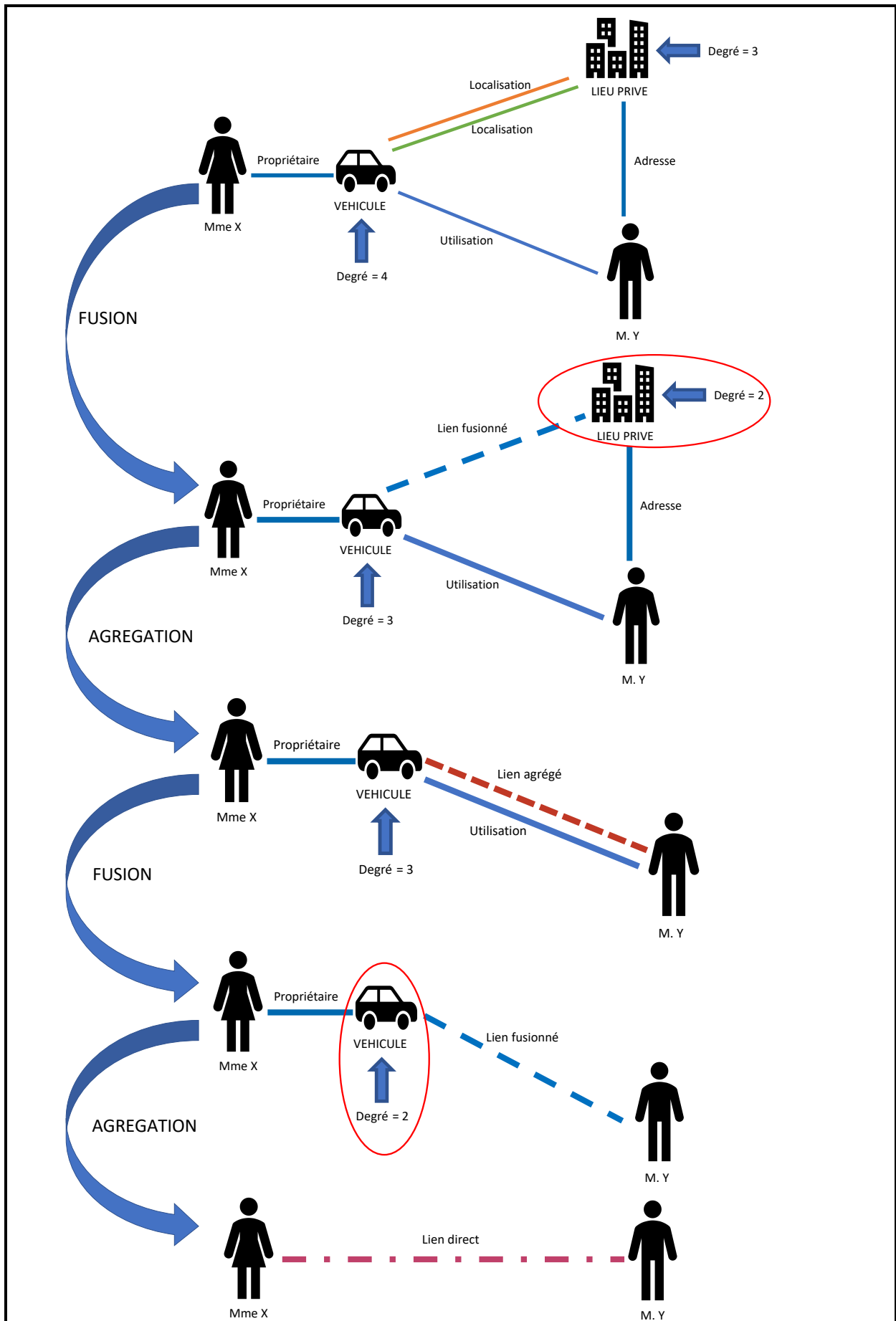


Figure 44 Processus itératif de fusion des arêtes en dérivation et d'agrégation des arêtes en série

Le principe de cette méthode réside dans le fait qu'à défaut d'extraire tous les chemins reliant chaque paire de personnes du graphe pour en étudier la sémantique, nous les analysons au cours des processus de réduction de la complexité, au moment où l'algorithme parcourt le graphe, en stockant les informations des trinôme initiaux dans chaque nouvelle arête créée, jusqu'à obtenir les liens directs contenant les chemins originaux.

Sur des données ainsi réduites, l'algorithme d'extraction de chemin peut alors travailler pour extraire les chemins n'ayant pas abouti au lien direct « personne-personne » recherché. Ces traitements sont décrits dans la section f) « Traitements des sous-graphes « personne-personne » *path_{ij}*: obtention des liens directs pour les cas les plus complexes ». Le tableau 37 indique les traitements effectués à chaque étape.

Tableau 37 Description de la hiérarchie de sous-graphes utilisée avec *Tulip* pour réduire la complexité du graphe *n-mode*

Sous-graphe	Graphe parent	Traitement	Nœuds	Arêtes
<i>data_root</i> (graphe initial)	/	Import des données ; Création des méta-nœuds pour les identités multiples ³⁴⁵	1 751	193 121
<i>data0</i>	<i>data_root</i>	Suppression des arêtes des trinômes dont la force liante est nulle ³⁴⁶ ; Remplacement des nœuds concernant des identités multiples par les méta-nœuds	1 743	2 345
<i>data1</i>	<i>data_root</i>	Fusion des arêtes multiples ³⁴⁷	1 743	2 100
<i>data2</i>	<i>data_root</i>	Suppression des composantes connexes contenant moins de deux personnes ; Suppression des nœuds de degré 1 qui ne sont pas de type « PERSONNE » ; Suppression des personnes isolées	1 099	1 470
<i>data3</i>	<i>data_root</i>	Création des méta-arêtes pour les nœuds de degré 2 qui ne sont pas de type « PERSONNE »	877	1 248

³⁴⁵ Réunion en méta-nœuds des sommets correspondants à une même personne sous plusieurs identités.

³⁴⁶ Voir section V.2 a) « Attribution des valeurs de force liante des symboles terminaux ».

³⁴⁷ *i.e.* suppression dans le sous-graphes des arêtes multiples pour les remplacer par les arêtes fusionnées.

Sous-graphe	Graphe parent	Traitement	Nœuds	Arêtes
<i>data4</i>	<i>data_root</i>	Fusion des arêtes multiples	877	1 189
<i>data5</i>	<i>data_root</i>	Suppression des nœuds de degré 1 qui ne sont pas de type « PERSONNE » ³⁴⁸ ; Suppression des personnes isolées	870	1 182
<i>data6</i>	<i>data_root</i>	Création des méta-arêtes pour les nœuds de degré 2 qui ne sont pas de type « PERSONNE »	836	1 148
<i>data7</i>	<i>data_root</i>	Fusion des arêtes multiples	836	1 140
<i>data8</i>	<i>data_root</i>	Suppression des nœuds de degré 1 qui ne sont pas de type « PERSONNE »	836	1 140
<i>data9</i>	<i>data_root</i>	Création des méta-arêtes pour les nœuds de degré 2 qui ne sont pas de type « PERSONNE »	832	1 135
<i>pathij</i>	<i>data_root</i>	Création des 12 996 sous-graphes correspondant aux 12 996 paires de personnes liées par des chemins irréductibles	832	16 955
<i>Social_network</i>	<i>data_root</i>	Création du réseau social obtenu avec tous les liens directs établis (dans <i>data9</i> ainsi que dans les graphes <i>personne-personne</i>).	665	12 996
<i>Investigation_network</i>	<i>data_root</i>	Filtrage des arêtes à conserver depuis <i>Social_network</i>	?	?

³⁴⁸ Nœuds initialement de degré > 1 devenus de longueur 1 suite à la fusion d'arêtes multiples.

f) Traitements des sous-graphes « personne-personne » $path_{ij}$: obtention des liens directs pour les cas les plus complexes

Le graphe réduit obtenu après avoir appliqué toutes les itérations des différentes opérations de réduction de la complexité, *data9*, comporte 527 nœuds dont 360 personnes, pour 646 arêtes (figure 45).

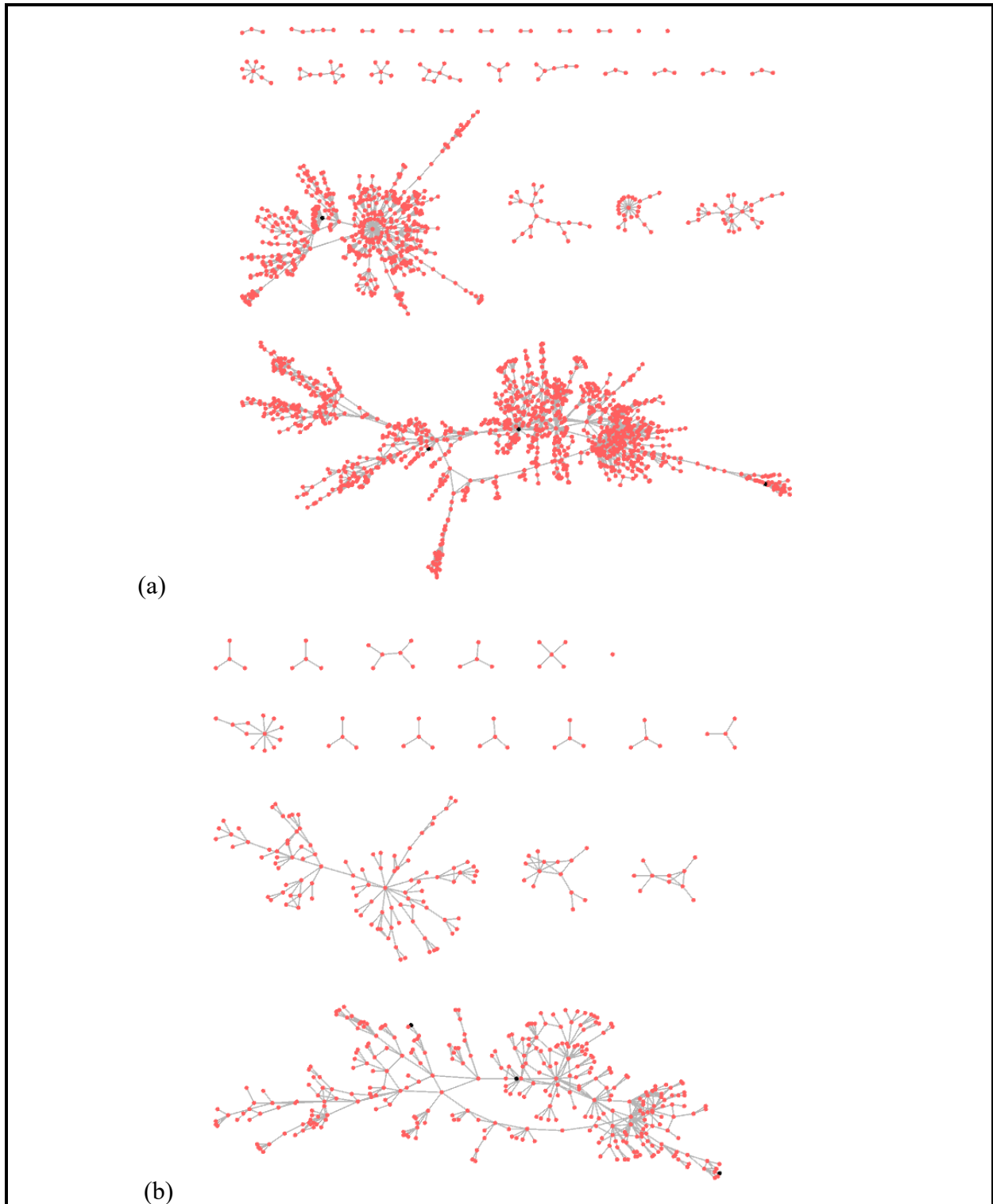


Figure 45 Graphe *data0* (a) et graphe réduit *data9* (b)

Parmi les 527 nœuds de *data9*, 360 sont de type « PERSONNE » et 167 représentent toujours plusieurs autres types d'entités. Les chemins restants ne peuvent être réduits d'avantage avec notre algorithme. En effet, toutes les arêtes multiples initiales et générées au cours de la réduction ont été fusionnées, ce qui signifie que le nombre d'arêtes, et par conséquent le degré de chaque nœud restant, ne peut plus diminuer. Or, la fonction d'agrégation ne s'applique qu'aux nœuds différents de « personne » et de degré 2, ce qui n'est le cas d'aucun des nœuds restants dans *data9*, comme l'illustrent les figures 46 et 47).

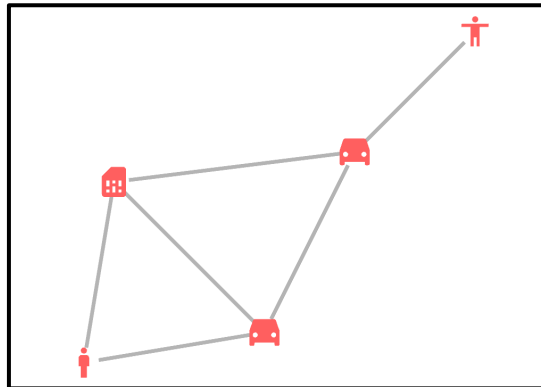


Figure 46 Exemple de chemin irréductible : les voitures et la carte SIM sont de degré 3, et il n'y a pas d'arêtes multiples.

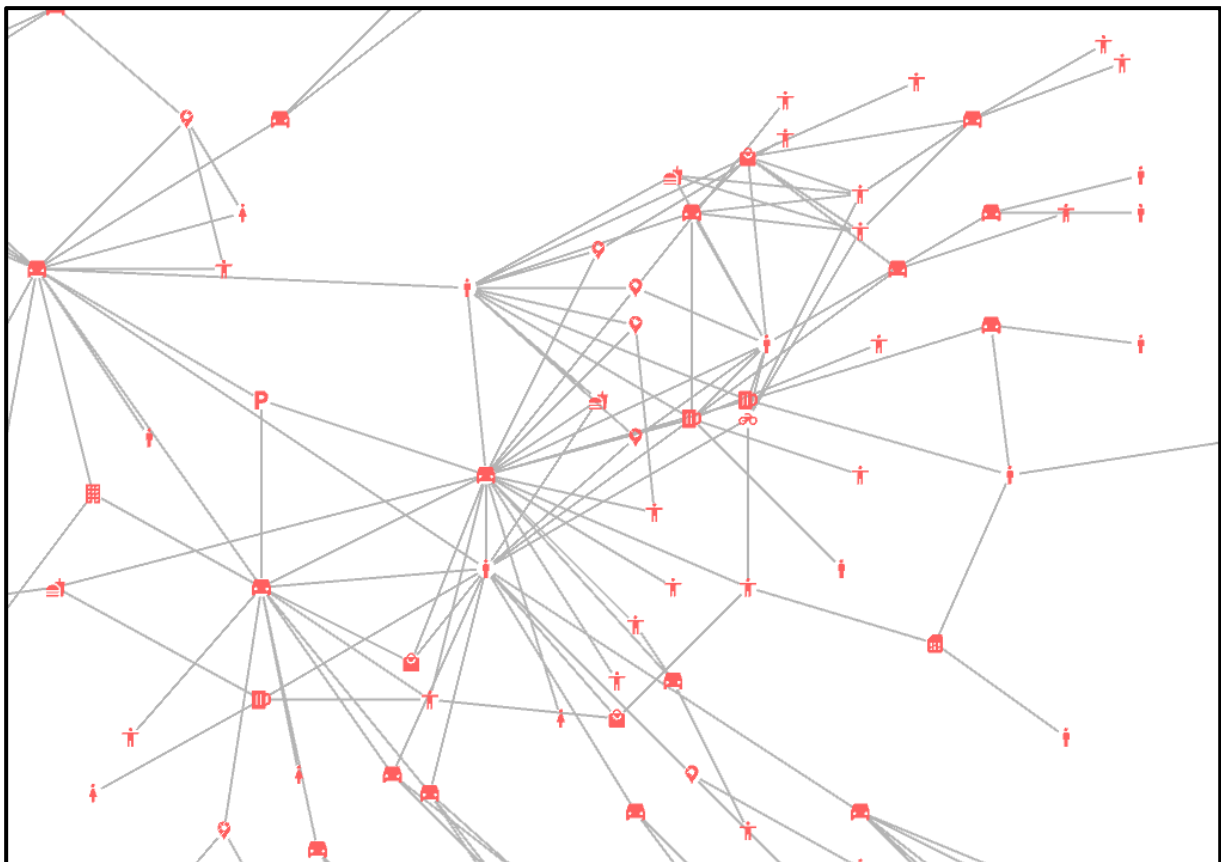


Figure 47 Extrait zoomé de *data9*

Cependant, le nombre d'arêtes et de nœuds a été suffisamment restreint pour qu'un ordinateur récent et relativement puissant soit capable de calculer en un temps raisonnable l'ensemble des chemins entre les 12 996 paires de personnes graphiquement connectées. Pour

chacune d'entre elles, un sous-graphe $path_{ij}$ contenant tous et uniquement les chemins liant les personnes i et j a été généré à partir de *data9*, soit 12 996 sous-graphes. L'algorithme de réduction de la complexité a alors été appliqué sur chacun des $path_{ij}$, permettant de réduire 8 420 sous-graphes à un lien direct unique pondéré entre les deux individus. 4 574 sous-graphes restent irréductibles, pour les mêmes raisons que précédemment (absence d'arêtes multiples et de nœuds de degré 2, cf. figure 48 et tableau 38).

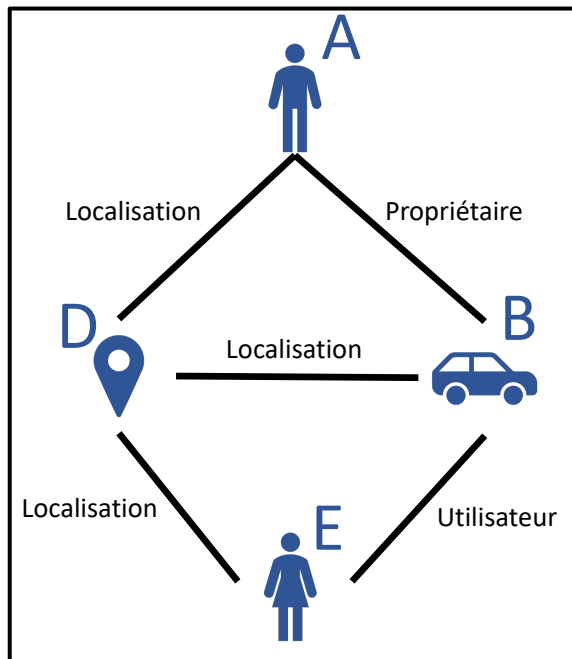


Figure 48 Exemple de sous-graphe irréductible

Tableau 38 Nombre d'arêtes des 4574 graphes irréductibles

Nombre d'arêtes	Nombre de graphes
5	15
6	435
7	2549
8	273
9	297
10	65
11	20
12	404
13	92
14	175
15	157
16	26
17	15
18	13
19	10
20	10
21	10
22	3
23	2
24	1
25	2
TOTAL	4574

Ces cas ont donc nécessité le développement de nouveaux traitements pour pouvoir calculer un lien direct unique pondéré. Dans le cas illustré en figure 48, quatre chemins permettent de relier les personnes A et E :

- *Chemin 1* : A-B-E
- *Chemin 2* : A-D-E
- *Chemin 3* : A-B-D-E
- *Chemin 4* : A-D-B-E

Fusionner ensemble ces chemins reviendrait à comptabiliser plusieurs fois des trinômes n'apparaissant pourtant qu'une seule fois³⁴⁹, ce qui biaiserait les résultats. Or, pour des raisons également énoncées précédemment³⁵⁰, il est important de considérer tous les chemins, et pas uniquement le plus court. En analysant la sémantique des chemins, il apparaît que ce type de configuration présente une redondance d'information :

- *Chemin 1* : A est propriétaire du véhicule B qui a été localisé à côté du véhicule C dont E est propriétaire ;
- *Chemin 2* : A a été localisé dans le parking D où a été localisé le véhicule C dont E est propriétaire ;
- *Chemin 3* : A est propriétaire du véhicule B localisé dans le parking D où a été localisé le véhicule C dont E est propriétaire ;
- *Chemin 4* : A a été localisé dans le parking D où a été localisé le véhicule B qui a été localisé à côté du véhicule C dont E est propriétaire.

Le fait que l'information soit répétée implique la possibilité de choisir les chemins afin de prendre en compte toutes les informations contenues dans les liens sans pour autant avoir de doublons. Nous avons fait le choix méthodologique de considérer qu'en présence d'arêtes multiples, la force liante maximale ne peut être diminuée par la présence de liens ayant une force liante inférieure³⁵¹. Il en va de même en présence de chemins multiples comme dans le cas présent. Ainsi, la valeur maximale de pondération du lien direct unique résultant a été recherchée. Cela consiste à calculer la meilleure combinaison de chemins garantissant que chaque arête ne soit comptabilisée qu'une seule fois.

Pour ce faire, dans un premier temps, les forces liantes de tous les chemins existants ont été calculées, puis ces derniers ont été classés par valeur décroissante. Le chemin ayant la force liante la plus élevée, C_{max} a été sélectionné. Ont été retirés parmi les chemins restants tous ceux ayant au moins une arête en commun avec C_{max} , puis le chemin ayant la deuxième meilleure pondération, C_{max2} , a été sélectionné. Le processus est répété jusqu'à ce que toutes les arêtes soient prises en compte une et une seule fois, avec pour conséquence que les chemins sélectionnés sont alors tous linéaires, c'est-à-dire que tous les nœuds intermédiaires sont de degré 2. Les fonctions d'agrégation puis de fusion ont alors été appliquées à ces chemins dans

³⁴⁹ {C-E} : commun aux quatre chemins ;

{A-B} : commun aux chemins 1 et 3 ;

{A-D} : commun aux chemins 2 et 4 ;

{B-C} : commun aux chemins 1 et 4 ;

{D-C} : commun aux chemins 2 et 3 ;

{B-D} : commun aux chemins 3 et 4 (puisque le graphe est non-orienté, {B-D est équivalent à {D-B}).

³⁵⁰ Cf section V.1 d) « Pré-traitements graphiques : réduction de l'espace de recherche ».

³⁵¹ Cf section V.1 b) « Règles de réécriture en série et en dérivation ».

les 4574 sous-graphes. Ainsi, pour chacune des paires de personnes graphiquement connectées dans *data0*, un lien direct, unique et pondéré a été obtenu.

Ainsi, dans l'exemple de la figure 48, le processus fonctionne comme suit :

1. Calcul des forces liantes des chemins :

- *Chemin 1* : $P(A-B-E) = 0,4875$
- *Chemin 2* : $P(A-D-E) = 0,4225$
- *Chemin 3* : $P(A-B-D-E) = 0,2925$
- *Chemin 4* : $P(A-D-B-E) = 0,2925$

Soit $P(A-B-E) > P(A-D-E) > P(A-B-D-E) = P(A-D-B-E)$

2. Sélection des chemins :

- Le chemin ayant la meilleure force liante est sélectionné : $C_{max} = \text{Chemin 1}$
- Les chemins ayant au moins une arête en commun avec C_{max} sont retirés, à savoir les *Chemin 3* et *4* (respectivement arêtes A-B et B-E)
- Le chemin restant, *Chemin 2* est également sélectionné : il ne partage aucune arête avec C_{max} , l'information n'est donc pas redondante et peut être conservée pour être également prise en compte.

3. Calcul du lien direct :

Le lien direct entre les personnes A et E est alors obtenu par fusion des chemins *1* et *2* sélectionnés.

Enfin, le graphe *social_network*, illustré en figures 49 et 50, a été généré : il contient toutes les paires de personnes connectées et les liens directs correspondants (calculés sur les sous-graphes *data0* à *data9* et les *path_{ij}*). Il possède douze composantes connexes, 665³⁵² nœuds, tous de type « PERSONNE » et 12996 arêtes (contre respectivement 16 composantes connexes, 1 743 entités dont 673 « PERSONNES » et 2 347 arêtes pour le graphe *n-mode data0*).

³⁵² Les personnes n'étant reliées à aucune autre personne n'étant pas ajoutées dans le réseau social.

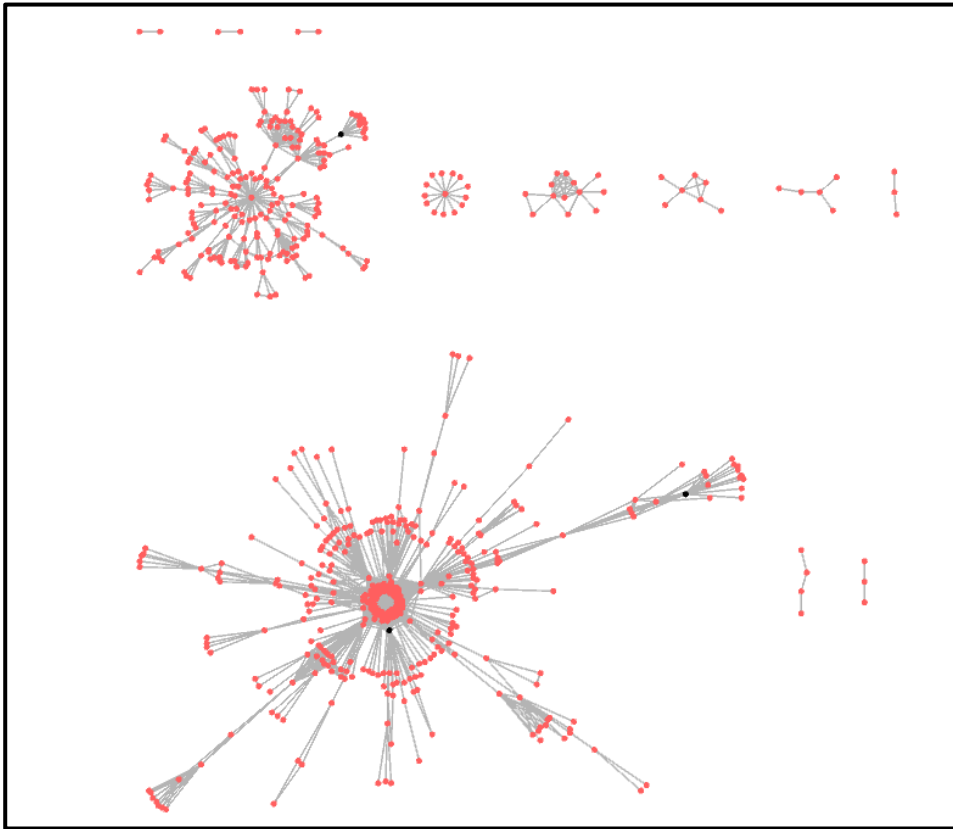


Figure 49 Réseau pondéré des personnes graphiquement connectées

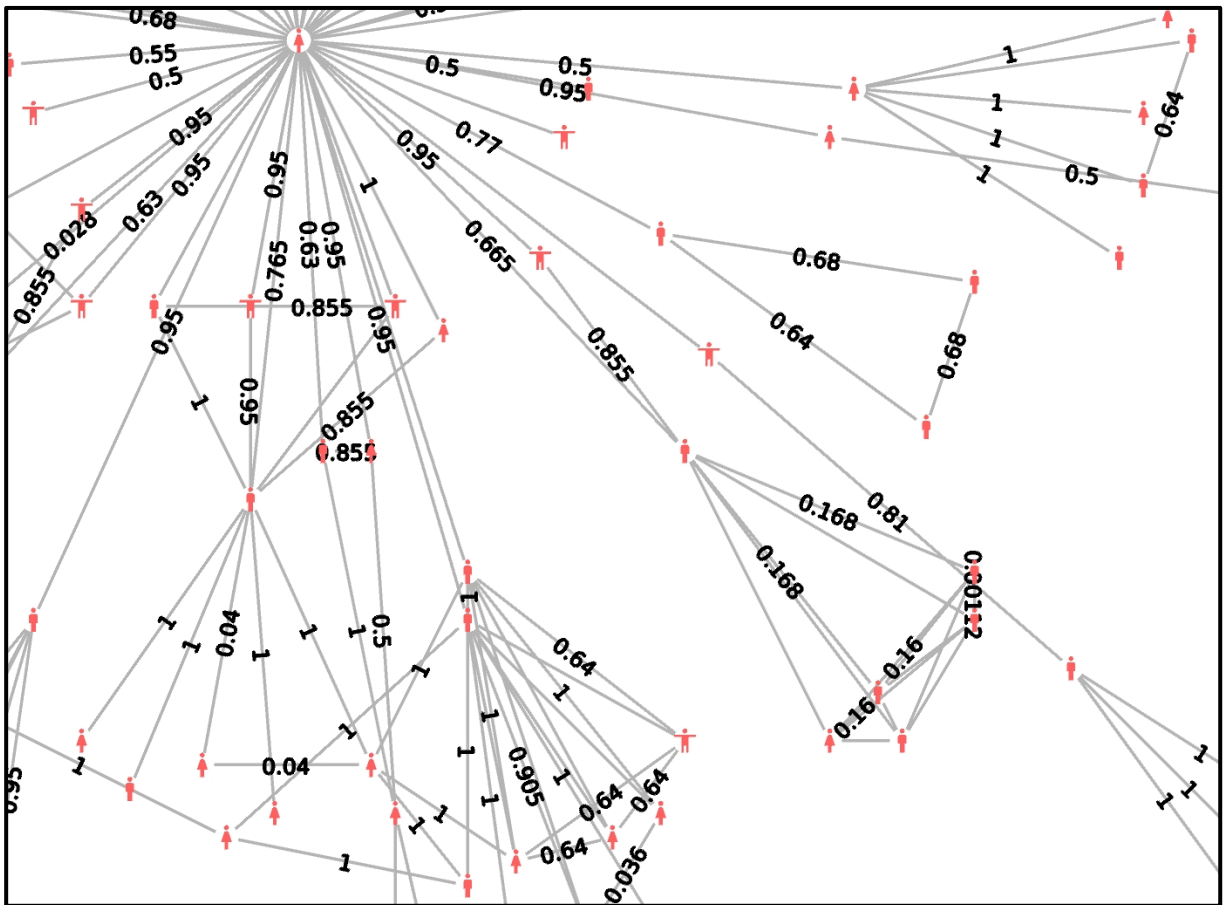


Figure 50 Réseau pondéré des personnes graphiquement connectées (zoom)

La dernière étape d'obtention du réseau social de l'organisation criminelle consistera à déterminer, sur la base des pondérations de ces arêtes, quelles connexions seront conservées et rejetées dans le sociogramme final³⁵³.

Les concepts méthodologiques étant posés, il s'agit à présent de paramétrer l'algorithme en définissant les valeurs des forces liantes à attribuer aux symboles terminaux puis à définir les fonctions de fusion et d'agrégation.

V.2 Paramétrage de l'algorithme d'extraction

a) Attribution des valeurs de force liante des symboles terminaux

Afin de proposer des résultats pertinents, la phase d'attribution des forces liantes des symboles terminaux (ou trinômes) est essentielle. En effet, de ces valeurs dérivent les pondérations finales des liens directs « personne-personne ». Ces pondérations seront le critère de sélection des liens directs à conserver : la décision finale d'accepter ou de rejeter l'existence d'une connexion entre deux personnes est prise selon la position de la pondération de leur lien direct par rapport à la valeur seuil S définie dans le chapitre VI.

Ainsi, dans un premier temps, un questionnaire à choix multiple a été mis en place et transmis à la communauté des 250 analystes criminels de la Gendarmerie nationale par le Lieutenant-Colonel Barlatier afin de tenir compte de leurs connaissances et de leur expérience au moment d'attribuer les valeurs (figure 51). Le questionnaire et les réponses sont disponibles en annexes 6 et 7. L'idée était de demander aux analystes, d'après leur expérience pratique, de qualifier et/ou quantifier³⁵⁴ le poids qu'ils accordent aux trinômes sémantiques³⁵⁵ lorsqu'ils évaluent si deux protagonistes d'une affaire sont en contact. L'échelle de qualification/quantification proposée est décrite dans le tableau 39.

³⁵³ Cf sections V.1 b) « Méthodologie de transformation : inspiration issue des concepts de grammaires formelles » et V.1 d) « Pré-traitements graphiques : réduction de l'espace de recherche ».

³⁵⁴ Selon la méthode de leur choix.

³⁵⁵ Par opposition aux trinômes brut basés sur les libellés, tels que {M. DUPONT – se gare à côté de – EPICERIE JOURDAIN}.

Tableau 39 Échelle de qualification/quantification des symboles terminaux³⁵⁶

Qualification	Quantification
Lien négligeable	0
Lien faible	2
Lien potentiellement fort	4
Lien fort	6
Lien très fort	8
Lien extrêmement fort	10

Une courte introduction expliquant l’objectif de la démarche, le cadre de l’affaire (trafic et blanchiment de fonds) et la méthodologie envisagée précédait le questionnaire. L’exercice a été difficile : il n’est pas évident d’expliquer en quelques lignes seulement qu’il s’agit de pondérer individuellement les trinômes tout en les évaluant dans des chemins complets, *i.e.* dans des chaînes de trinômes successif composant des chemins entre des personnes (*cf.* figure 28 – Analogie explicative de la notion de force liante).

1 Dans le cadre d'une enquête sur un GCO, quel poids attribueriez-vous aux types de liens suivants ?

Propositions de réponse: *Choisissez une seule réponse dans chaque rangée*

	0 - lien négligeable	2 - lien faible	4 - lien potentiellement fort	6 - lien fort	8 - lien très fort	10 - lien extrêmement fort
PERSONNE -> adresse -> LIEU (caractère public/privé non précisé)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PERSONNE -> localisation -> LIEU PRIVE	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PERSONNE -> localisation -> LIEU PUBLIC	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PERSONNE -> localisation -> LIEU (caractère public/privé non précisé)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PERSONNE -> propriétaire -> LIEU (caractère public/privé non précisé)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PERSONNE MORALE -> adresse -> LIEU (caractère public/privé non précisé)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 51 Illustration de la première des sept questions soumises aux analystes criminels de la GN

En outre, il existe dans les données 69 symboles terminaux ; il n’était pas envisageable de tous les soumettre au questionnaire – qui aurait alors demandé à plus de 100 analystes d’y passer un temps considérable au détriment de leur travail – il a donc été nécessaire de sélectionner au préalable un nombre restreint de trinômes.

³⁵⁶ Comme indiqué précédemment, les forces liantes seront comprises entre 0 et 1. Un changement d’échelle (de 0 à 10) a été effectué dans le questionnaire afin de rendre l’attribution des pondérations par les analystes plus tangible.

Pour ce faire, les trinômes sémantiques existants dans les données à l'issue des pré-traitements ont été recensés, ainsi que leurs occurrences. Le tableau 40 présente les 27 symboles terminaux mentionnés dans le questionnaire, regroupés sur sept questions³⁵⁷. La liste complète des trinômes sémantiques est disponible en annexe 5. Dix-sept des 18 symboles terminaux les plus fréquents ont été choisis, ainsi que ceux concernant le lien entre deux entités de type « PERSONNE » et ceux dont la force de la sémantique était particulièrement délicate à quantifier dans le contexte spécifique d'une enquête criminelle³⁵⁸.

³⁵⁷ La huitième et dernière question concerne d'éventuels commentaires, remarques, critiques et conseils de la part des analystes criminels.

³⁵⁸ À titre d'exemple le trinôme {PERSONNE-localisation-LIEU PUBLIC}, dont la force liante serait relativement insignifiante dans un cadre quotidien, mais pourrait s'avérer élevée dans le contexte de gérants commerces de façade couvrant du blanchiment d'argent.

Tableau 40 Liste, occurrences et fréquence des 27 symboles terminaux soumis au questionnaire

Symbole terminal	Occurrences	Fréquence
PERSONNE-adresse-LIEU (ADRESSE)	293	12,43%
PERSONNE-utilisateur-VÉHICULE	261	11,07%
PERSONNE-titulaire-VÉHICULE	177	7,51%
VÉHICULE-localisation-LIEU (ADRESSE)	169	7,17%
PERSONNE-familial amical-PERSONNE	162	6,87%
PERSONNE-en contact-PERSONNE	155	6,58%
PERSONNE-utilisateur-TÉLÉCOM	137	5,81%
PERSONNE MORALE-adresse-LIEU (ADRESSE)	97	4,12%
PERSONNE-échange objet ou argent-PERSONNE	92	3,90%
PERSONNE-professionnel-PERSONNE MORALE	92	3,90%
PERSONNE-localisation-LIEU PUBLIC	81	3,44%
TÉLÉCOM-en contact-TÉLÉCOM	65	2,76%
PERSONNE-localisation-LIEU (ADRESSE)	55	2,33%
VÉHICULE-localisation-LIEU PUBLIC	48	2,04%
PERSONNE-criminel-PERSONNE	35	1,48%
PERSONNE MORALE-titulaire-VÉHICULE	33	1,40%
PERSONNE-est-PERSONNE	31	1,32%
PERSONNE-titulaire-LIEU (ADRESSE)	18	0,76%
PERSONNE-échange objet ou argent-VÉHICULE	18	0,76%
VÉHICULE-localisation-VÉHICULE	15	0,64%
PERSONNE-fait partie-PERSONNE	6	0,25%
PERSONNE-professionnel-PERSONNE	5	0,21%
PERSONNE-adresse-PERSONNE	4	0,17%
PERSONNE-échange objet ou argent-LIEU (ADRESSE)	4	0,17%
PERSONNE-échange objet ou argent-LIEU PUBLIC	4	0,17%
PERSONNE-utilisateur-PERSONNE	3	0,13%
PERSONNE-localisation-LIEU PRIVÉ	1	0,04%

L'intégralité des résultats est disponible en annexe 7. L'analyse des critiques du questionnaire et de la méthode envisagée, de la variabilité des valeurs attribuées aux différents trinômes par un même analyste et d'un analyste à l'autre ainsi que des temps de réponses a mis en évidence que l'exercice de réponse a été aussi difficile pour les analystes que la formulation des questions ne l'a été pour la doctorante. Elle a aussi mis en évidence les divergences de points de vue sur la question³⁵⁹. Sur les 250 analystes contactés, 128 s'y sont prêtés, ce qui constitue un échantillon suffisant pour prendre en compte les réponses en ayant une vision

³⁵⁹ Les réponses au questionnaire étant anonymes, il n'a pas été possible de déterminer si les opinions variaient simplement d'un gendarme à l'autre, ou si ces dernières dépendaient de critères précis (tels que le grade, le niveau d'expérience, la zone d'action, la spécialité...)

globale. Les résultats ont été analysés symbole par symbole (tableau 41), et une attention particulière a été apportée aux remarques, commentaires et critiques des répondants (question 8).

Tableau 41 Résultats des réponses à la première des sept questions soumises aux analystes criminels de la GN

1 Dans le cadre d'une enquête sur un GCO, quel poids attribueriez-vous aux types de liens suivants ?

Matrice de choix uniques, Nombre de répondants 128 x, sans réponse 0 x

Réponse	● 0 - lien négligeable	● 2 - lien faible	● 4 - lien potentiellement fort	● 6 - lien fort	● 8 - lien très fort	● 10 - lien extrêmement fort
PERSONNE -> adresse -> LIEU (caractère public/privé non précisé)	3 (2,3%)	24 (18,8%)	40 (31,3%)	24 (18,8%)	17 (13,3%)	20 (15,6%)
PERSONNE -> localisation -> LIEU PRIVE	1 (0,8%)	8 (6,3%)	37 (28,9%)	30 (23,4%)	26 (20,3%)	26 (20,3%)
PERSONNE -> localisation -> LIEU PUBLIC	6 (4,7%)	42 (32,8%)	49 (38,3%)	13 (10,2%)	12 (9,4%)	6 (4,7%)
PERSONNE -> localisation -> LIEU (caractère public/privé non précisé)	9 (7,0%)	30 (23,4%)	52 (40,6%)	13 (10,2%)	12 (9,4%)	12 (9,4%)
PERSONNE -> propriétaire -> LIEU (caractère public/privé non précisé)	5 (3,9%)	9 (7,0%)	25 (19,5%)	30 (23,4%)	28 (21,9%)	31 (24,2%)
PERSONNE MORALE -> adresse -> LIEU (caractère public/privé non précisé)	5 (3,9%)	37 (28,9%)	44 (34,4%)	9 (7,0%)	20 (15,6%)	13 (10,2%)

Pour chaque question, les réponses sont extrêmement diversifiées. Or, dans le présent cas de figure, il n'aurait pas été pertinent de baser le choix des valeurs des forces liantes sur les indicateurs statistiques classiquement utilisés, tels que la valeur majoritaire, la moyenne ou la médiane. En effet, utiliser, pour chaque symbole, la valeur la plus attribuée par les répondants aurait impliqué un biais considérable, dans la mesure où chacun des choix possibles comptabilise globalement 10% à 20% des réponses. Le taux maximal a été obtenu pour le trinôme {PERSONNE – localisation – LIEU (ADRESSE)} auquel 40,6% des répondants ont affecté une valeur comprise entre 4 et 6, correspondant au qualificatif « lien potentiellement fort ». Si ce taux peut sembler élevé, attribuer automatiquement cette pondération à ce trinôme aurait impliqué de ne pas tenir compte de 59,4% des réponses qui proposaient une autre valeur, proportion trop importante statistiquement pour être ignorée. De même, valider la médiane aurait consisté à sélectionner la valeur au-delà et en deçà de laquelle se situaient 50% des réponses, rejetant ainsi la grande majorité des points de vue des analystes. Enfin, quel que soit l'indicateur, il renvoie une valeur comprise entre 4 et 6 pour l'ensemble des symboles terminaux soumis, comme l'illustrent les figures 52 à 54. Or, il est nécessaire d'avoir des valeurs

diversifiées sur l'ensemble des symboles terminaux pour les discriminer selon leur force liante : choisir l'un de ces trois indicateurs statistiques comme la valeur de force liante à attribuer n'aurait pas été cohérent avec ce besoin.

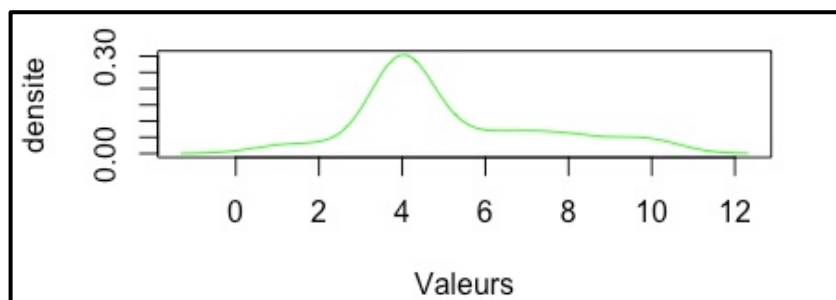


Figure 52 Distribution des valeurs majoritaires des réponses sur l'ensemble des 27 symboles terminaux

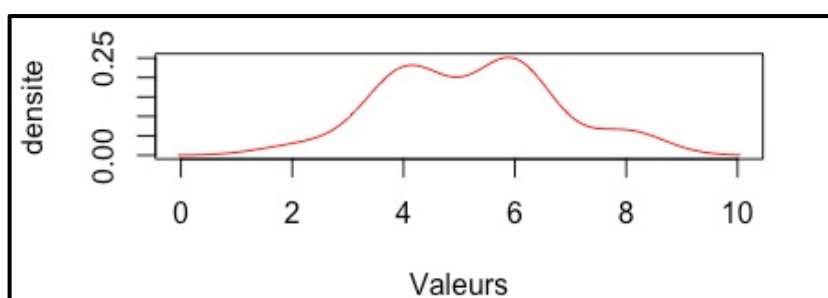


Figure 53 Distribution des moyennes des valeurs-réponses sur l'ensemble des 27 symboles terminaux soumis

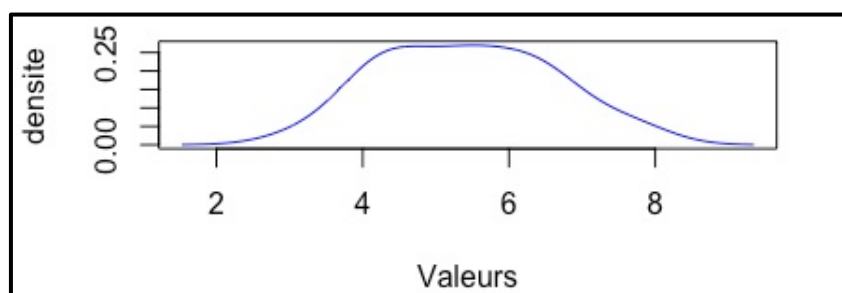


Figure 54 Distribution des médianes des valeurs-réponses sur l'ensemble des 27 symboles terminaux soumis

Finalement, des critères ont été mis en place pour classer les symboles terminaux selon une hiérarchie sémantique afin d'évaluer leurs forces liantes de manière relative. En effet, plus que les valeurs absolues des pondérations, le classement des trinômes les uns par rapport aux autres selon ces valeurs est important : il doit refléter le raisonnement d'une intelligence humaine, qui accorderait naturellement plus de confiance à l'existence d'un lien entre deux individus localisés au même moment dans un lieu privé (tel qu'une maison) que dans un lieu public (tel qu'un centre commercial).

Les résultats du questionnaire, et en particulier les commentaires et remarques, ont été utilisés comme une base solide pour guider l'évaluation relative des symboles terminaux. Cependant, cognitivement parlant, il est extrêmement difficile de positionner 69 éléments sur une même échelle de valeur. Les trinômes de signification proche ont donc été regroupés afin

de pouvoir comparer à l'intérieur d'un même groupe un nombre réduit d'éléments. Les groupes peuvent correspondre soit à un type d'entité source et d'entité destination communs à tous ses éléments, soit à un type d'entité source et de lien associé commun à tous ses éléments.

En sciences criminelles, lors de l'analyse des indices physiques, sont distingués les liens directs, permettant de lier une trace à une personne (tels que l'ADN ou les traces digitales) des liens indirects, qui lient une trace à un objet qui pourra (ou non) être lié à une personne (tels qu'une trace de semelle, de pneu, les fibres d'un vêtement)³⁶⁰. Cette distinction est ici reprise pour évaluer les groupes : un premier critère discrimine les trinômes concernant directement un individu des autres.

Le second critère oppose les trinômes se référant à des événements ponctuels (tels qu'une localisation) et ceux référant à des événements inscrits dans la durée (tels qu'une adresse). Les 13 groupes sont référencés dans le tableau 42 et les trinômes appartenant aux groupes 2 et 3 sont énumérés dans le tableau 43 à titre d'exemple. La liste des symboles terminaux et de leur groupe d'appartenance est détaillée dans l'annexe 8.

³⁶⁰ Alexandre Girod, Christophe Champod et Olivier Ribaux, *Traces de souliers*, Lausanne, Presses polytechniques et universitaires romandes, 2008.

Tableau 42 Groupes sémantiques constitués pour l'attribution des forces liantes

Groupe	Base	Sémantique
1	Typologie des entités source et destination	PERSONNE - ... - PERSONNE
2	Typologie des entités source et destination	PERSONNE - ... - LIEU
3	Typologie des entités source et destination	PERSONNE (PHYSIQUE/MORALE) - ... - OBJETS
4	Typologie d'entité source – typologie du lien	TÉLÉCOM- en contact - ...
5	Typologie d'entité source – typologie du lien	PERSONNE- localisation - ...
6	Typologie d'entité source – typologie du lien	VÉHICULE- localisation - ...
7	Typologie d'entité source – typologie du lien	PERSONNE- échange objet ou argent - ...
8	Typologie d'entité source – typologie du lien	PERSONNE MORALE - ... - LIEU
9	Typologie d'entité source – typologie du lien	OBJETS - localisation - ...
10	Typologie d'entité source – typologie du lien	PERSONNE (PHYSIQUE/MORALE) - professionnel - ...
11	Typologie d'entité source – typologie du lien	OBJETS - liens - ...
12	Typologie d'entité source – typologie du lien	PERSONNE - ... - OBJETS
13	Typologie d'entité source – typologie du lien	Inclassable

Tableau 43 Énumération des symboles terminaux classés dans les groupes 2 et 3

Symbole terminal	Groupe sémantique
PERSONNE-adresse-LIEU (ADRESSE)	2
PERSONNE-adresse-LIEU PRIVÉ	2
PERSONNE-adresse-LIEU PUBLIC	2
PERSONNE-adresse-PERSONNE MORALE	2
PERSONNE-titulaire-LIEU (ADRESSE)	2
PERSONNE-utilisateur-LIEU PRIVÉ	2
PERSONNE MORALE-titulaire-TÉLÉCOM	3
PERSONNE-titulaire-TÉLÉCOM	3
PERSONNE-utilisateur-TÉLÉCOM	3
PERSONNE-utilisateur-VÉHICULE	3
PERSONNE-titulaire-VÉHICULE	3
PERSONNE MORALE-titulaire-VÉHICULE	3
PERSONNE MORALE-utilisateur-VÉHICULE	3
PERSONNE-titulaire-OBJET	3
LIEU PUBLIC-titulaire-VÉHICULE	3
LIEU PUBLIC-titulaire-TÉLÉCOM	3

Une troisième étape a consisté à attribuer à chaque groupe un ordre de grandeur de la force liante et à en vérifier la cohérence relative inter-groupes. Les valeurs de force liante sont bornées et appartiennent à l'intervalle [0 ; 1] ; 0 signifiant que la sémantique du trinôme ne fournit pas d'élément incitant à considérer que le lien direct existe³⁶¹, 1 signifiant que le lien est

³⁶¹ Il ne s'agit pas ici d'affirmer qu'une telle connexion n'existe pas dans la réalité, mais de dire que, si elle existe, l'information à disposition ne permet pas de la mettre en évidence.

avéré. Outre les résultats et les commentaires issus du questionnaire, le contexte dans lequel ont été collectées et saisies les données a été pris en considération. En effet, il ne s'agit pas d'informations relationnelles entre des entités prises au hasard récoltées aléatoirement dans une population générale, mais de données recueillies dans le cadre d'une enquête de gendarmerie et ayant fait l'objet d'une préanalyse puis d'une sélection par l'analyste criminel lors de leur saisie dans la matrice. La préanalyse est une étape permettant, par des méthodes simples de confrontation, structuration et représentation de la donnée, de lui donner du sens, d'effectuer un premier tri de sa pertinence ainsi que d'orienter le recueil de nouvelles informations³⁶².

En quatrième étape, les forces liantes ont enfin été attribuées à chaque symbole terminal en contrôlant la cohérence relative intra-groupe. Au-delà du contexte, la forme de criminalité concernée par l'enquête a également été prise en compte, à savoir le trafic et blanchiment de fonds. En effet, comme l'ont très justement souligné certains analystes en commentaire du questionnaire, la force liante des trinômes n'est pas nécessairement généralisable à toutes les activités criminelles : « Bonjour, la classification des liens varie en fonction du type de GCO³⁶³ travaillé. Pour un GCO étranger qui sévit en France puis quitte le territoire, les alias seront très importants et les liens TÉLÉCOM potentiellement forts. Alors que pour la communauté des GDV³⁶⁴, les liens téléphoniques n'auront pas grande importance et les alias seront souvent inexistant³⁶⁵ ». Cinquièmement, pour certains symboles terminaux, tels que {PERSONNE – autre – LIEU (ADRESSE)}, essentiellement issus des groupes 10 à 13 et ayant des fréquences inférieures à 1%, il a été nécessaire d'aller rechercher les dénominations des entités et les libellés des liens correspondants dans les données pour évaluer la force liante à attribuer.

Pour les symboles terminaux non soumis au questionnaire, bien que les différentes remarques aient fourni un cadre de réflexion, l'expérience des analystes n'a pas pu être recueillie. Ainsi, les forces liantes attribuées à ces trinômes ont été contrôlées et corrigées par le Lieutenant-Colonel Barlatier.

Ce processus a nécessité plusieurs semaines de réflexion, essais et échecs avant d'être abouti. Une étape de validation des forces liantes a été effectuée « à froid », c'est-à-dire après quatorze jours, en classant cette fois les symboles terminaux non par groupe mais par valeur de

³⁶² Jérôme Barlatier, cours « Collecte de la donnée », stage d'analyse en renseignement criminel, SCRC, 2018.

³⁶³ Groupe criminel organisé.

³⁶⁴ Gens Du Voyage.

³⁶⁵ Retranscription d'une remarque d'un répondant (anonyme).

force liante décroissante, s'assurant ainsi de leur cohérence relative générale. La liste des symboles et de leurs pondérations est disponible en annexe 8.

Précisons que, pour des raisons techniques, de temps de calcul de l'algorithme, les données ont été importées en considérant les liens comme non-dirigés, *i.e.* indifféremment orientés de l'entité source dans les données à l'entité destination ; ce qui a eu pour conséquence la génération d'un graphe (*data0*) dit « non-dirigé », c'est-à-dire dans lequel les arêtes n'ont pas de sens directionnel. En d'autres termes, les trinômes {M. DUPONT – se gare à côté de – EPICERIE JOURDAIN} et {EPICERIE JOURDAIN – se gare à côté de – M. DUPONT} sont considérés comme équivalents, ce qui signifie que l'algorithme peut parcourir le symbole terminal correspondant indifféremment dans le sens initial {PERSONNE – localisation – LIEU PUBLIC} que dans le sens inverse {LIEU PUBLIC – localisation – PERSONNE}. Aussi, lorsque les symboles terminaux ont été extraits, nous avons veillé à attribuer aux trinômes sémantiques inverses les mêmes forces liantes (tableau 44).

Tableau 44 Exemple de valeurs de force liante

Symbole terminal (et inverse)	Groupe sémantique	Force liante
PERSONNE- échange objet ou argent -PERSONNE (ou PERSONNE- échange objet ou argent-PERSONNE)	1	1
PERSONNE-utilisateur-VÉHICULE (ou VÉHICULE-utilisateur-PERSONNE)	3	0,65
PERSONNE-titulaire-VÉHICULE (ou VÉHICULE-titulaire-PERSONNE)	3	0,75
PERSONNE MORALE-titulaire-VÉHICULE (ou VÉHICULE-titulaire-PERSONNE MORALE)	3	0,1
TÉLÉCOM-en contact-TÉLÉCOM	4	0,9
PERSONNE-localisation-LIEU PUBLIC (ou LIEU PUBLIC-localisation-PERSONNE)	5	0,1
VÉHICULE-localisation-LIEU PUBLIC (ou LIEU PUBLIC-localisation-VÉHICULE)	6	0,08
VÉHICULE-localisation-LIEU PRIVÉ (ou LIEU PRIVÉ-localisation-VÉHICULE)	6	0,6
PERSONNE-échange objet ou argent-VÉHICULE (ou VÉHICULE- échange objet ou argent -PERSONNE)	7	0,65
PERSONNE-professionnel-LIEU PUBLIC (ou LIEU PUBLIC-professionnel-PERSONNE)	10	0,4

Il est important de rappeler que ces pondérations chiffrées reflètent une valeur sémantique : dans la mesure où il n'est pas possible d'appliquer des opérations de fusion et d'agrégation directement sur les symboles terminaux, la démarche consiste à évaluer le niveau de confiance induit par un chemin complet dans l'existence d'une connexion entre ces personnes source et destination en transitant par des valeurs et des opérations mathématiques. Ces pondérations, au même titre que les fonctions de fusion et d'agrégation décrites dans la section suivante, sont donc un moyen d'imiter un raisonnement humain basé sur des mots et non une fin arithmétique en elle-même.

Enfin, pour tenir compte de la solidité des liens, c'est-à-dire de leur caractère confirmé ou hypothétique³⁶⁶, la force liante des liens non confirmés est systématiquement divisée par deux.

³⁶⁶ Cf. section III.2 c) 3 « Relations décrites et informations associées ».

b) Règles de réécriture en série et en dérivation

La méthode établie dans les précédentes sections de ce chapitre consiste, pour chaque paire d'individus graphiquement reliés par au moins un chemin, à générer un lien unique direct pondéré entre les personnes source et destination. Pour ce faire, nous avons développé un algorithme qui, en partant des chemins graphiques initiaux, va fusionner les arêtes multiples (liens en dérivation) existantes entre deux entités et agréger entre eux les symboles terminaux composant chaque chemin (liens en série) pour former le symbole terminal recherché, à savoir {PERSONNE – en contact – PERSONNE}. Tous ces chemins graphiques ne permettant pas de conclure si les individus source et destination ont effectivement une connexion directe, nous proposons de quantifier le niveau de confiance en l'existence de ces potentielles relations par une analyse formelle de la sémantique des symboles terminaux qui en composent les chemins. Dans la mesure où il n'est pas possible d'appliquer les opérations de fusion et d'agrégation directement sur les symboles, nous avons attribué à chacun d'entre eux une pondération reflétant sa force liante d'un point de vue sémantique. Ces valeurs sont ensuite fusionnées et agrégées pour calculer la pondération du symbole {PERSONNE – en contact – PERSONNE} correspondant à chaque chemin.

Ainsi, nous cherchons à obtenir un résultat sémantique, soit une évaluation de la signification verbale de chaque chemin. Les fonctions de fusion et d'agrégation doivent respecter ce principe, c'est-à-dire qu'elles sont définies sur la base de conditions qui ne sont pas d'ordre mathématique.

Dans un premier temps, une revue de littérature a été effectuée afin de caractériser l'évolution de la force de la relation entre deux entités en fonction d'une part du nombre de liens les connectant (arêtes multiples) et d'autre part de la distance les séparant (liens en série). Cependant, dans le cas présent, il ne s'agit pas de déterminer la force d'un lien avéré, mais d'inférer sur son existence en se basant sur la « force liante » des intermédiaires qui le composent. La décision a donc été prise de poser les deux principes suivants, relativement instinctifs : premièrement, plus le nombre d'arêtes reliant deux nœuds donnés est important, plus la force liante du trinôme générique {ENTITE – lien fusionné – ENTITE} est élevée ; deuxièmement, à sémantique équivalente, la confiance dans l'existence d'une connexion directe entre deux personnes diminue lorsque la distance graphique les séparant augmente.

Afin de respecter ce premier principe, la fonction de fusion doit être croissante : la valeur de sortie, *i.e.* celle de la force liante du lien fusionné doit augmenter avec le nombre (et les valeurs) des forces liantes des symboles terminaux qui le composent. Ceci implique que lorsque des liens multiples de forces liantes fortes et faibles sont fusionnés, la valeur de l'arête résultante

ne peut être inférieure à la force liante maximale des liens multiples initiaux³⁶⁷. Cependant, dans la mesure où les fusions et agrégations sont effectuées de manière itérative, il est impératif que la valeur de sortie de la fonction reste dans l'intervalle [0 ;1]. En effet, un lien fusionné n'étant pas nécessairement un résultat final, mais bien souvent un résultat intermédiaire amené à être à nouveau fusionné et/ou agrégé avec d'autres trinômes, il est important pour conserver la cohérence que leurs valeurs respectives restent du même ordre de grandeur.

Enfin, il est essentiel que la fonction prenne en compte dans ses paramètres le fait que les pondérations ne sont pas linéaires, c'est-à-dire que neuf occurrences d'un lien de force liant égale à 0,1 seront toujours, d'un point de vue sémantique, moins impactantes qu'une seule occurrence d'un lien de force liante égale à 0,9. À titre d'exemple, le fait qu'une personne soit localisée à neuf reprises dans un lieu public (*i.e.* neuf occurrences du trinôme {PERSONNE – localisation – LIEU PUBLIC} dont la p force liante vaut 0,1 n'est pas équivalent au fait qu'une personne utilise une ligne téléphonique trinôme {PERSONNE – utilisateur – TÉLÉCOM} dont la force liante vaut 0,9). La fonction de fusion doit donc obéir aux conditions suivantes : être croissante, non linéaire et retourner une valeur appartenant à l'intervalle [0 ; 1].

Afin de répondre à ce critère de non-linéarité sans négliger ni la croissance ni les bornes de la fonction, les symboles terminaux ont, dans un premier temps, été répartis en trois classes de force liante (sur une base sémantique), matérialisées par trois intervalles (tableau 45).

Tableau 45 Classes de force liante des liens et équivalence en pondération

Classe	Qualification	Intervalle
<i>C1</i>	Force liante élevée	[1 ; 0,6]
<i>C2</i>	Force liante moyenne]0,6 ; 0,1]
<i>C3</i>	Force liante faible]0,1 ; 0]

Dans un second temps, la fonction de fusion a été définie de la manière suivante : la force liante du lien fusionné sera égale à la force liante maximale des liens initiaux à laquelle s'ajoute 0,1 par lien de classe *C1* ; 0,05 par lien de classe *C2* et 0,01 par lien de classe *C3*. Nous voulons cependant rester conforme à la classification sémantique établie, *i.e.* le fait que malgré leur multiplicité, des liens faiblement liants ne résultent pas en une trop importante force liante globale. Les valeurs obtenues sont donc bornées selon les classes auxquelles appartiennent les

³⁶⁷ Ce choix méthodologique peut être discuté dans la mesure où, dans des arêtes multiples, il serait envisageable que certains types de liens viennent diminuer la force liante des autres arêtes avec lesquels ils ont fusionnés (tels que des liens impliquant une absence ou une négation). Dans le cas présent, cette possibilité ne pourrait s'appliquer qu'à deux symboles, qui n'ont chacun qu'une seule occurrence sur les 2 357 liens étudiés : {PERSONNE – non lien – PERSONNE} et {PERSONNE – lien interdit – PERSONNE}. Compte tenu de leurs occurrences négligeables et de la complexité déjà élevée du modèle, la décision a été prise d'utiliser une fonction de fusion croissante et de contourner le biais en attribuant à ces deux symboles terminaux une force liante nulle.

liens multiples initiaux. Ainsi, s'ils contiennent uniquement des liens de classe $C3$, la force liante maximale qui peut être obtenue est 0,15 ; s'ils contiennent des liens de classe $C2$ (ou de classes $C2$ et $C3$), la force liante maximale obtenue est de 0,65 et s'ils contiennent au moins un lien de classe $C1$, la valeur maximale est de 0,95. Ainsi, le changement de classe lors du passage des liens multiples au lien unique fusionné est possible si le nombre de liens initiaux est suffisamment important, la force liante du lien fusionné étant toutefois restreinte à une valeur basse de la classe supérieure. Enfin, dans la mesure où la valeur 1 correspond à une connexion établie et confirmée entre deux individus, et où ce niveau de certitude ne peut être rigoureusement obtenu par le biais de cette méthode, la force liante maximale calculée est bornée à 0,95.

Mathématiquement, cela se synthétise comme suit :

Soient :

- n_{C1} le nombre de liens de classe 1,
- n_{C2} le nombre de liens de classe 2,
- n_{C3} le nombre de liens de classe 3,
- n le nombre de liens total,
- p_i les pondérations (forces liantes) des arêtes, avec $p_i \in [0 ; 1]$ et $i \in [1 ; n]$

$$f_{fusion} : p_i \rightarrow P_{arête\ fusionnée} = f_{fusion}(p_i)$$

Si $\{p_i\} \subset \{C3\}$ alors :

$$P_{arête\ fusionnée} = \min(0.15 ; \max(p_i) + 0.01 \cdot (n_{C3} - 1))$$

Sinon, si $\{p_i\} \subset \{C2\} \cup \{C3\}$ alors :

$$P_{arête\ fusionnée} = \min(0.65 ; \max(p_i) + 0.05 \cdot (n_{C2} - 1) + 0.01 \cdot n_{C3})$$

Sinon, i.e. $\{p_i\} \subset \{C1\} \cup \{C2\} \cup \{C3\}$ alors :

$$P_{arête\ fusionnée} = \min(0.95 ; \max(p_i) + 0.1 \cdot (n_{C1} - 1) + 0.05 \cdot n_{C2} + 0.01 \cdot n_{C3})$$

La fonction d'agrégation a pour seules contraintes d'être bornée entre $[0 ; 1]$ et d'être décroissante : plus un chemin contient d'intermédiaires, plus la valeur de la pondération du lien direct « personne-personne » associé sera faible. Ainsi, lorsque deux trinômes sont agrégés en

une méta-arête, la fonction d'agrégation consiste simplement à multiplier entre elles leurs forces liantes :

Soient p_1 et p_2 les pondérations des deux arêtes à agréger, avec p_1 et $p_2 \in [0 ; 1]$:

$$f_{agrégation} : (p_1 ; p_2) \rightarrow P_{méta-arête} = f_{agrégation}(p_1 ; p_2)$$

$$P_{méta-arête} = p_1 \cdot p_2$$

Il est à noter que les fonctions f_{fusion} et $f_{agrégation}$ sont toutes les deux commutatives et associatives, c'est-à-dire que ni l'ordre des symboles minimaux dans le chemin ni l'ordre dans lequel sont effectuées les opérations de fusion et d'agrégation n'influe sur le résultat final, à savoir la valeur des pondérations des liens directs uniques générés dans *Social_network*³⁶⁸.

La méthodologie globale est schématisée figure 55.

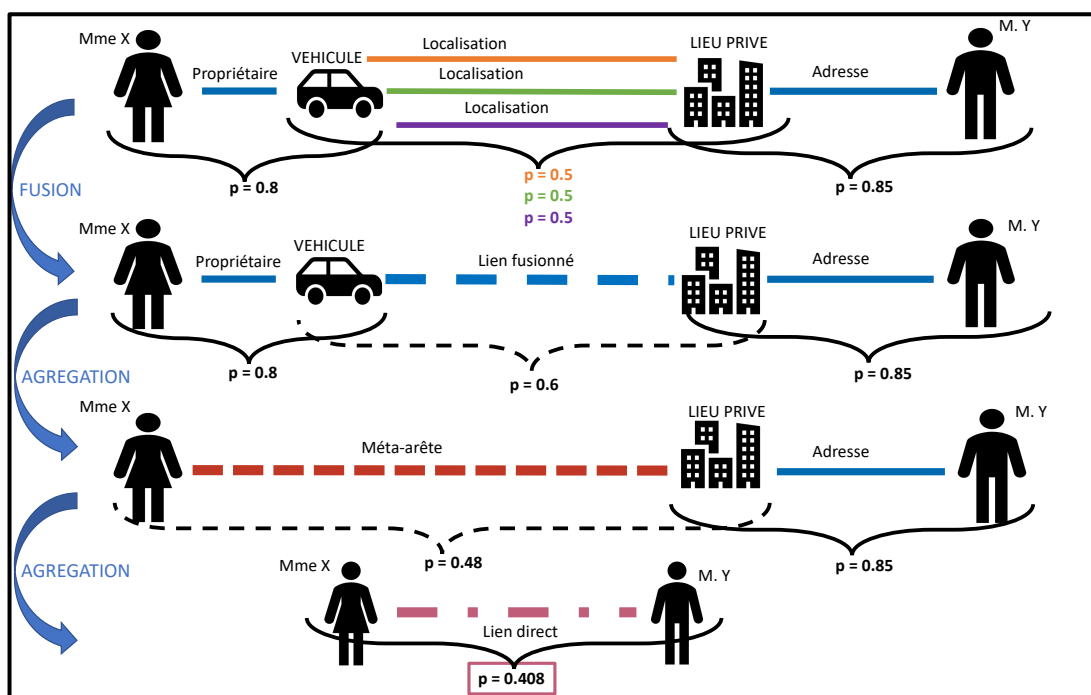


Figure 55 Schéma de résumé de la méthode de pondération des liens directs entre individus

c) Le réseau social pondéré *Social_network*

La méthode décrite dans les sections précédentes a ainsi permis d'obtenir un réseau social, représenté par le graphe *Social_network*, à partir du graphe d'enquête multivarié

³⁶⁸ Bruno Pinaud, Maud Benichou et Guy Melançon, « Extraction d'un réseau social criminel par transformation d'un graphe d'enquête multivarié », *op. cit.*. Article accepté pour publication en format long par le comité de lecture.

data0 (figure 56). Il s'agit d'un graphe simple, c'est-à-dire sans boucle (arête récurrente sur un seul nœud) ni arêtes multiples.

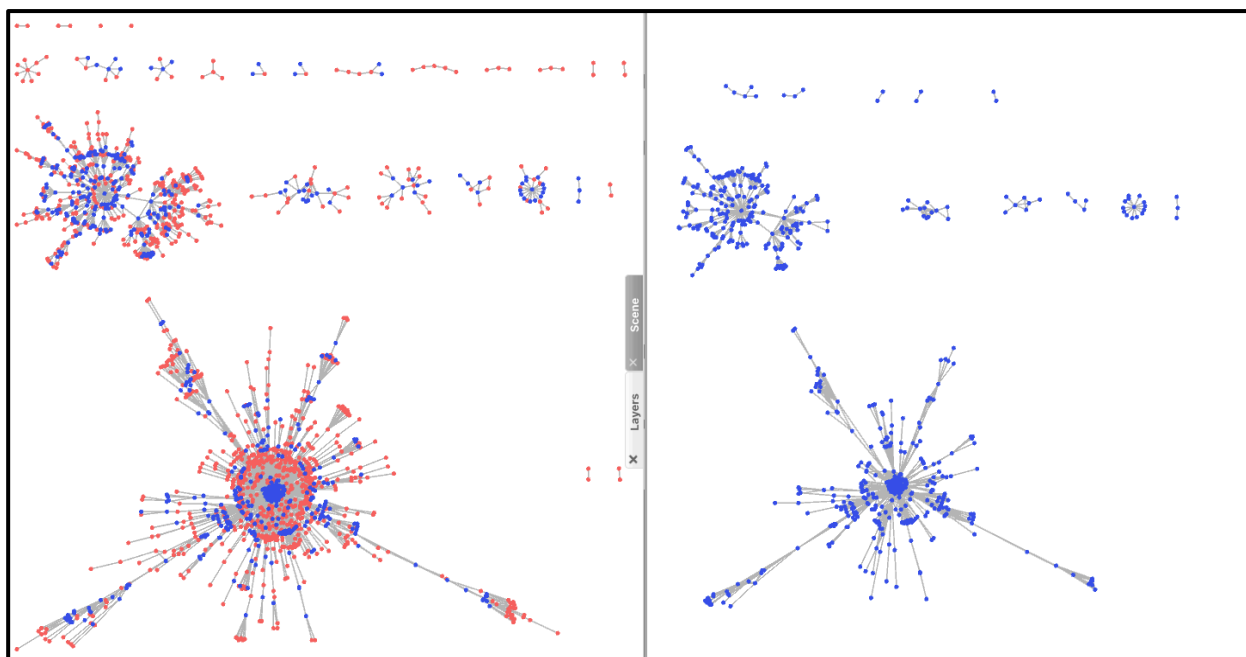


Figure 56 Graphes *data0* (à gauche) et *Social_network* (à droite). Les personnes sont représentées par les points bleus.

Social_network contient 665 sommets, tous de type « PERSONNE » et 12 996 arêtes pondérées par leur force liante (figure 57 et 58).

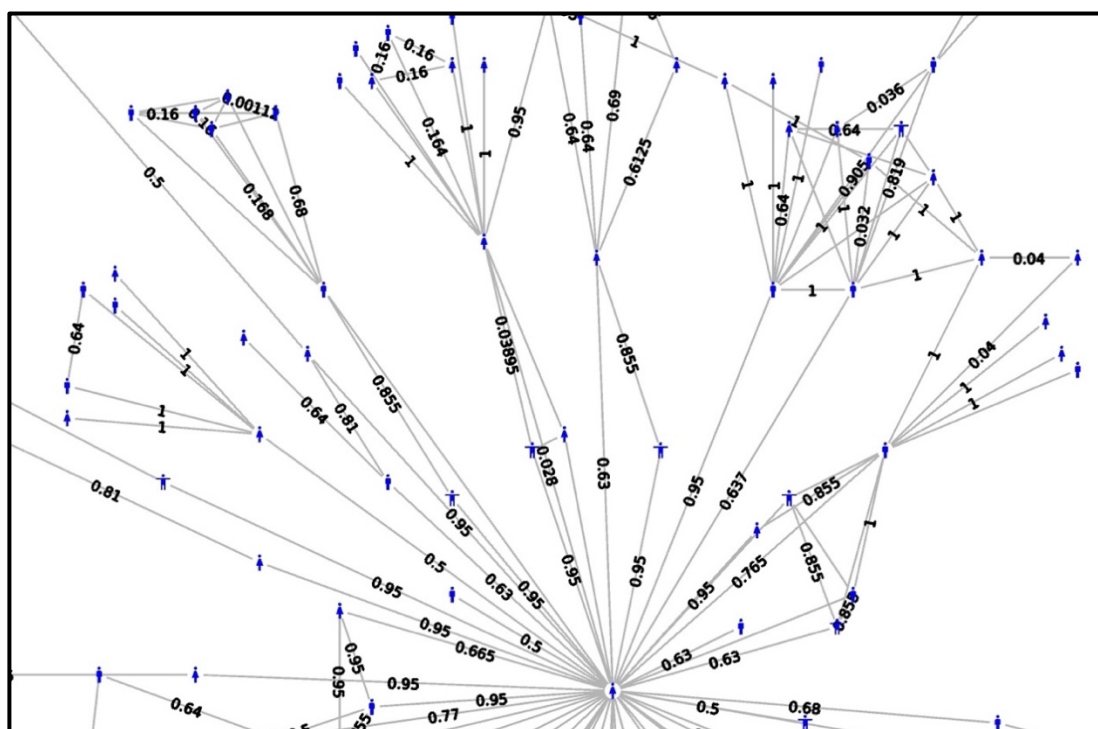


Figure 57 Zoom sur *Social_network* avec affichage des pondérations

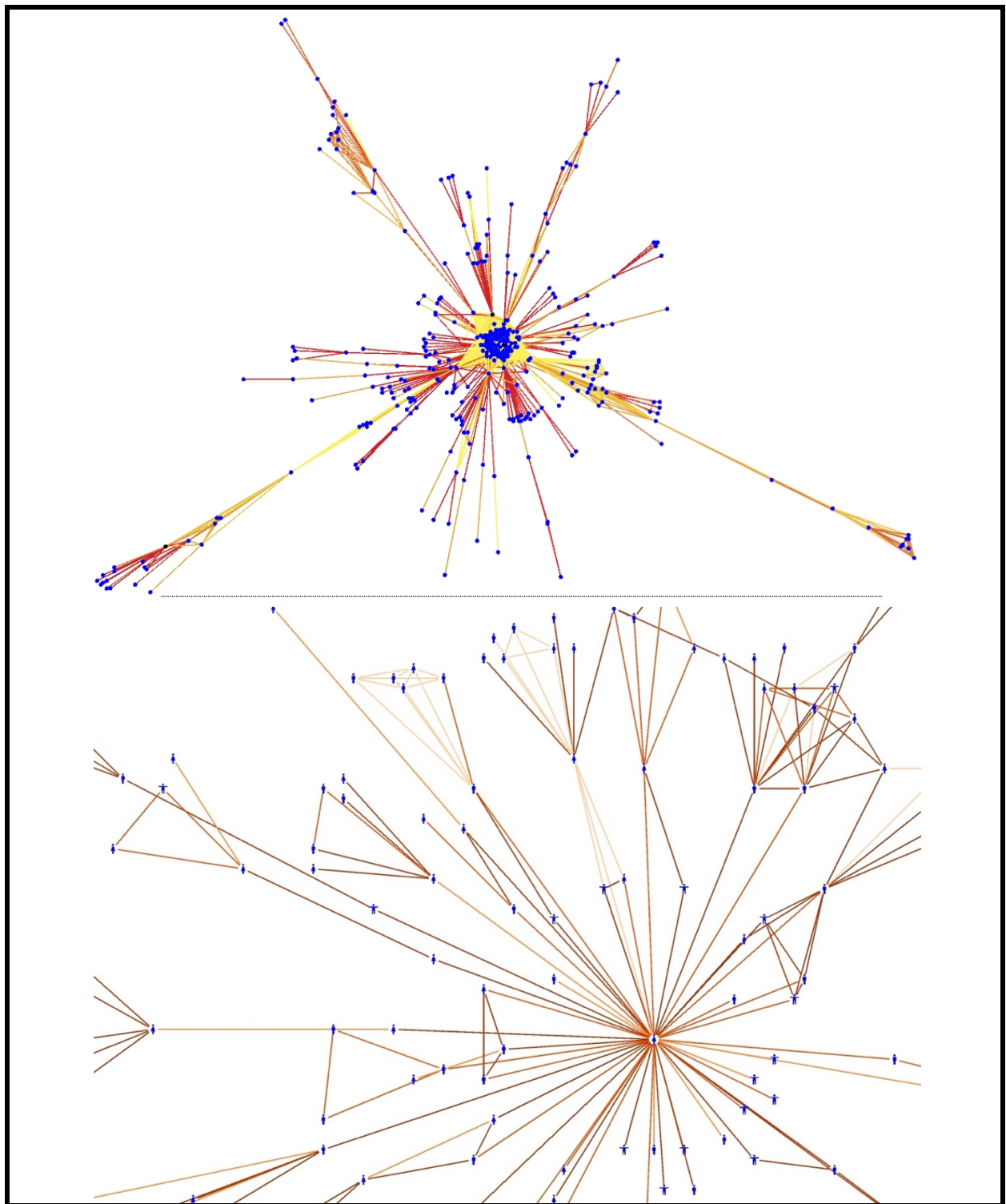


Figure 58 Deux niveaux de zoom sur *Social_network* avec une échelle de couleur sur les pondérations. Les liens les plus clairs ont les plus faibles pondérations, les plus foncés ont les plus élevées.

La force liante de chaque arête reflète la vraisemblance sur le terrain du lien représenté dans *Social_network*. Ainsi, nous avons à ce stade obtenu un réseau social fictif, dans la mesure où, parmi tous les liens entre individus induits depuis le graphe d'enquête *data0* dans le réseau social *Social_network*, un certain nombre ne correspond pas à une connexion réelle entre les deux individus sur le terrain. En effet, les liens dans *Social_network* ont été établis lorsqu'il existait dans *data0* au moins un chemin graphique reliant deux individus donnés. Cependant,

comme nous l'avons vu, la sémantique de ces chemins ne permet pas nécessairement de conclure à l'existence d'un lien direct entre les individus source et destination. Il est donc nécessaire de filtrer les arêtes de *Social_network* pour ne conserver que celles qui correspondent, sur une base sémantique, à un lien vraisemblable : un seuil S d'acceptation ou rejet des liens doit être déterminé. Seules les arêtes dont la force liante est supérieure ou égale à S seront conservées dans le sociogramme de l'enquête, *i.e.* le graphe représentant le réseau social des personnes impliquées, *Investigation_network*.

Le chapitre VI expose dans une première section l'analyse des résultats obtenus par la méthode d'extraction présentée. Dans une deuxième section est décrite la détermination du seuil de filtrage S à appliquer au graphe *Social_network* pour obtenir le sociogramme de l'enquête. En troisième section est discutée, à la lueur de l'analyse des résultats, la méthode d'extraction et de filtrage du réseau social.

La méthodologie proposée fait l'objet d'un article accepté pour publication en format long par le comité de lecture pour la conférence « Extraction et gestion de connaissances (EGC) »³⁶⁹ qui se déroulera à Lyon du 16 au 20 janvier 2023.

³⁶⁹ <https://egc2023.sciencesconf.org/>

CHAPITRE VI DISCUSSION DE LA MÉTHODE D'EXTRACTION ET DÉTERMINATION DU SEUIL DE FILTRAGE

VI.1 Analyse des résultats de l'extraction du graphe social *Social_network* depuis le graphe multivarié *data0*

Nous disposons à présent d'un réseau social pondéré extrait du graphe d'enquête multivarié initial.

a) Analyse des pondérations des liens du graphe *Social_network*

Le graphe *Social_network* ainsi obtenu comporte 12 composantes connexes, 665 personnes, 12 996 relations toutes pondérées par leur force liante calculée lors de l'extraction. Son diamètre est de 8 (contre un diamètre de 23 pour le graphe initial *data0*.), ce qui signifie que la distance géodésique entre les deux individus les plus éloignés est de 7 (respectivement 22) personnes (respectivement entités) intermédiaires. La distribution des pondérations des arêtes a été étudiée et est illustrée en figure 59.

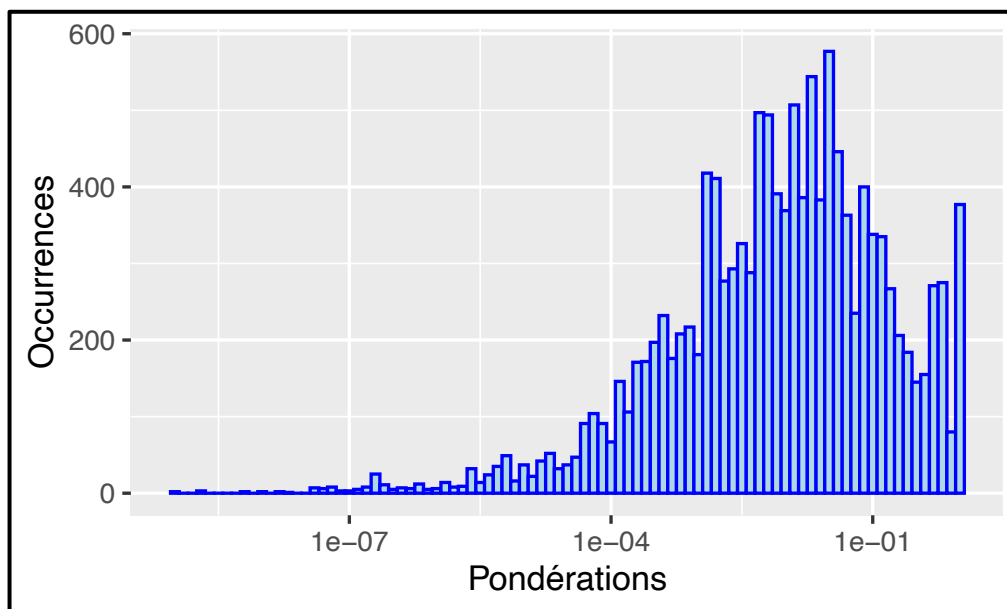


Figure 59 Distribution des pondérations des arêtes de *Social_network* (échelle x en \log_{10})

Les forces liantes calculées reflètent des résultats sémantiques et permettent de quantifier dans quelle mesure une chaîne sémantique permet (ou non) de considérer l'existence d'un lien entre deux individus. Afin d'une part d'estimer la performance du modèle et d'autre

part de déterminer un seuil de filtrage, un échantillon de chemins dans *data0* a été sélectionné manuellement et aléatoirement. Cet échantillon contient, pour des longueurs de chemins allant de 2 à 6 liens intermédiaires, 5 chemins par longueur, soit 25 chemins au total.

Il s'agit d'un échantillon extrêmement réduit. Ce choix a été fait pour plusieurs raisons. D'une part, si l'algorithme a été conçu pour historiser les informations à chaque étape, c'est-à-dire pour conserver les symboles contenus dans les chemins initiaux de *data0* lors de la génération des arêtes de *Social_network*, il ne permet pas à l'heure actuelle d'extraire les *patterns* de chemins de manière automatisée dans un format facilement exploitable. Il permet, en revanche, si l'analyste s'intéresse à une arête donnée *e* de *Social_network*, de remonter manuellement au(x) chemin(s) dont elle est dérivée. En effet, l'extraction de l'ensemble des chemins n'étant pas techniquement envisageable³⁷⁰, nous nous sommes concentrés lors de la conception de l'algorithme sur l'élaboration de la méthodologie – non triviale – de réduction de l'espace de recherche et de la complexité pour pouvoir traiter tous les chemins. En d'autres termes, l'information est effectivement disponible dans *Tulip*, mais son exploitation n'est pas évidente. Il s'agit là d'une des limites du modèle, discutée dans la section VI.3.

Ainsi, pour la création de l'échantillon, les chemins ont dû être extraits manuellement uns par uns en « se promenant » dans le graphe *data0*. D'autre part, pour estimer la validité du modèle, l'analyse de la sémantique des *patterns* doit également être effectuée manuellement afin de décider si les chemins doivent être acceptés ou rejetés³⁷¹. Or, les données contiennent plus de 2 750³⁷² *patterns* de chemins différents, il n'est donc pas envisageable de tous les analyser à la main. Ainsi, pour effectuer les analyses manuelles dans un délai raisonnable, le nombre de 5 chemins par longueur a été choisi pour l'échantillon.

Enfin, concernant la longueur maximale de chemins à analyser manuellement, la comparaison des résultats obtenus pour d'une part des *patterns* de chemins de longueur 2 à 5 et d'autre part des *patterns* de chemins de longueur 2 à 6 a montré que les chemins de longueurs 6 ne modifiaient pas les résultats³⁷³, raison pour laquelle les *patterns* de longueur supérieure à 6 n'ont pas été évalués. Cette démarche est cohérente avec les calculs de puissance statistiques effectués notamment dans les études épidémiologiques, qui consistent à calculer la taille minimale de l'échantillon de population nécessaire à la robustesse souhaitée de l'étude

³⁷⁰ Cf. section CHAPITRE V d) « Pré-traitements graphiques : réduction de l'espace de recherche ».

³⁷¹ Comme le font actuellement les analystes criminels et les analystes en renseignement criminel.

³⁷² Lors des premiers essais de conception et construction de l'algorithme, nous avons tenté d'extraire les chemins personne-personne, mais le script est tombé en erreur après avoir extrait 1 048 576 chemins, dont 7 253 *patterns* distincts.

³⁷³ Cf. annexe 9.

réalisée³⁷⁴. Ici, la puissance du test ne varie pas que l'on considère ou non les chemins de longueur 6. Enfin, les 5 chemins de longueur 6 ont été rejetés, avec une pondération maximale de l'ordre 10^{-3} , ce qui est déjà faible par rapport aux pondérations des *patterns* de longueur inférieure, et donc au seuil envisagé.

Ainsi, l'analyse sémantique a été effectuée manuellement sur les 25 chemins de l'échantillon provenant de *data0*³⁷⁵. Les résultats, c'est-à-dire la décision d'accepter ou de rejeter les liens induits par les *patterns* étudiés, ont ensuite été comparés aux forces liantes correspondant à ces *patterns*. Les *patterns* étudiés sont disponibles en annexe 9 ; le tableau 46 résume les résultats.

Tableau 46 Décision sémantique et pondération des 25 *patterns* de chemins étudiés

Longueur	N° pattern	Décision	Pondération
6	L6.5	Rejet	4,33E-04
6	L6.4	Rejet	2,70E-05
6	L6.3	Rejet	2,56E-04
6	L6.2	Rejet	5,83E-03
6	L6.1	Rejet	1,11E-03
5	L5.5	Rejet	1,92E-04
5	L5.4	Rejet	3,84E-05
5	L5.3	Acceptation	1,05E-01
5	L5.2	Rejet	2,34E-04
5	L5.1	Rejet	5,07E-02
4	L4.5	Acceptation	2,70E-03
4	L4.4	Rejet	8,32E-02
4	L4.3	Rejet	6,56E-01
4	L4.2	Acceptation	3,24E-02
4	L4.1	Rejet	4,23E-03
3	L3.5	Acceptation	4,10E-01
3	L3.4	Rejet	1,20E-02
3	L3.3	Acceptation	5,20E-03
3	L3.2	Rejet	2,60E-03
3	L3.1	Acceptation	7,29E-01
2	L2.5	Acceptation	4,23E-01
2	L2.4	Rejet	1,60E-03
2	L2.3	Acceptation	8,10E-01
2	L2.2	Acceptation	6,40E-01
2	L2.1	Acceptation	4,85E-01

³⁷⁴ Jacob Cohen, *Statistical power analysis for the behavioral sciences*, Hillsdale, N.J, L. Erlbaum Associates, 1988, 567 p ; Michel Cucherat, Michel Lièvre, Alain Leizorovicz et Jean-Pierre Boissel, *Lecture critique et interprétation des résultats des essais cliniques pour la pratique médicale*, Paris, Flammarion médecine-sciences, 2004 ; R. Murray Lindsay, « Publication System Biases Associated with the Statistical Testing Paradigm* », *Contemporary Accounting Research*, vol. 11, n° 1, 1994, pp. 33-57.

³⁷⁵ Seuls ont été considérés des chemins entièrement linéaires, c'est-à-dire sans liens multiples. Pour plus de rigueur, une évaluation similaire devrait être effectuée en considérant des chemins contenant des liens multiples.

Avant même de pousser les analyses plus loin, des corrections nécessaires au bon fonctionnement du modèle ont été repérées. Ces dernières sont de deux ordres : les informations manquantes et l'optimisation des paramètres.

Les chemins de l'échantillon dont l'analyse a mis en évidence des anomalies ou a soulevé d'importantes questions sont répertoriés dans le tableau 47 ci-dessous. Un code couleur nous permettra de nous référer aux éléments concernés lors de l'exposé des correctifs à apporter.

Tableau 47 *Patterns mettant en évidence la nécessité d'améliorer le modèle*

N°	Pattern	Décision sémantique	Pondération
L3.2	PERSONNE-utilisateur-VÉHICULE-titulaire-PERSONNE MORALE-professionnel-PERSONNE	Rejet	2,60E-03
L3.3	PERSONNE-localisation-LIEU pu-localisation-VÉHICULE- utilisateur-PERSONNE	Acceptation ?	5,20E-03
L3.4	PERSONNE-titulaire-LIEU ad-adresse-PERSONNE MORALE-professionnel-PERSONNE	Rejet ?	1,20E-02
L4.1	PERSONNE-utilisateur-VÉHICULE-titulaire-PERSONNE MORALE-titulaire-VÉHICULE-utilisateur -PERSONNE	Rejet	4,23E-03
L4.2	PERSONNE-titulaire-TEL-en contact-TEL-titulaire-LIEU pu- professionnel-PERSONNE	Acceptation ?	3,24E-02
L4.3	PERSONNE- titulaire-TEL-en contact-TEL-en contact-TEL- titulaire-PERSONNE	Rejet	6,56E-01
L4.5	PERSONNE-utilisateur-VÉHICULE-localisation-LIEU pu- localisation-VÉHICULE-utilisateur-PERSONNE	Acceptation ?	2,70E-03

Légende :
En orange : les chaînes problématiques,
En bleu : les entités à mieux définir,
En violet : les symboles dont la pondération doit être corrigée,
En vert : les symboles pour lesquels les attributs du lien doivent être pris en compte dans le modèle.

b) Optimisation des paramètres

Le tout premier paramètre du modèle est la catégorisation, à l'aide d'ontologies, des différents liens et entités. Deux entités en particulier nécessiteraient d'être mieux catégorisées : les « LIEUX PUBLICS » et les « PERSONNES MORALES », en bleu dans le tableau 47. En effet, l'analyse des chemins L4.2 et L3.3 a directement révélé que la catégorie « LIEU PUBLIC » n'est pas assez précise, dans la mesure où elle regroupe tous les commerces

accueillant du public indifféremment de leur taille, des petits commerces aux grands centres commerciaux. L'analyse des données originales a montré que 24 des 51 entités classées comme « LIEU PUBLIC », correspondent à des chaînes aux effectifs conséquents (hôtels, stations-service...), tandis que les 27 autres sont des commerces indépendants (épiceries, cafés...). Il en va de même pour les entités de type « PERSONNE MORALE », pour lesquelles, pour plus de précision dans l'attribution de la pondération, il faudrait avoir une estimation de la taille de l'entreprise, car les forces liantes des trinômes contenant cette entité seront très différentes s'il s'agit d'une multinationale ou d'une PME. En effet, notre objectif est de faire (ou non) des liens entre des personnes, et les interactions sociales ne sont pas de même nature dans ces différents cas.

En outre, il serait intéressant de savoir si les enquêteurs et analystes ont des raisons de penser que certaines entités sont en lien avec l'affaire. En effet, la force liante d'un lieu est sensiblement différente lorsque deux individus s'y retrouvent pour échanger de l'argent, que lorsqu'ils viennent y déposer de l'argent.

Pour plus de précision, il faudrait créer sur ces bases davantage de catégories distinctes, et effectuer à nouveau l'extraction et l'analyse. Cela n'est pas réalisable dans le délai imparti par la thèse³⁷⁶, aussi, pour les besoins de la recherche, nous devons, pour ces entités, fixer ces critères puis attribuer les forces liantes adéquates et refaire tourner l'algorithme d'extraction³⁷⁷ avant d'être en mesure de décider si les liens directs induits par les chemins contenant des « LIEUX PUBLICS » et « PERSONNES MORALES » doivent être acceptés ou rejetés. La distinction entre ces deux catégories est rappelée dans le tableau 48.

³⁷⁶ En effet, la création des ontologies est la toute première étape, et comme indiqué dans le chapitre IV, un grand nombre de paramétrages découle de cette dernière. Modifier l'ontologie à ce stade de la recherche impliquerait de recommencer l'ensemble des processus de paramétrage (ce qui n'est plus faisable) avant de pouvoir recommencer l'extraction.

³⁷⁷ Dans la mesure où l'algorithme réalise l'extraction en un peu moins de 20 minutes, ce n'est pas le fait d'exécuter à nouveau les scripts qui est problématique, mais bien de réappliquer l'ensemble de la méthode pour déterminer les nouveaux paramètres. Ainsi, modifier le fichier de pondération pour effectuer à nouveau les calculs reste faisable à ce stade, contrairement à la modification du nombre et de la sémantique des symboles minimaux.

Tableau 48 Catégorisation des entités comme « LIEUX PUBLICS » ou « PERSONNES MORALES »

Type de personne morale	Icône	CATEGORIE
Bureaux, locaux et sièges sociaux d'entreprises (n'accueillant pas de public)	Bureau	PERSONNE MORALE
Bars et cafés	Café	LIEU PUBLIC
Boutiques, supermarchés, centres commerciaux, hôtels	Magasin	LIEU PUBLIC
Restaurants et <i>fast-food</i>	Restaurant	LIEU PUBLIC

1. Pour les « LIEUX PUBLICS »

Les commerces, cafés, hôtels et restaurants, qui accueillent du public, ont été catégorisés comme des lieux, tandis que les bureaux et locaux ont été catégorisés comme des entreprises.

Afin de déterminer comment considérer dans le modèle les « LIEUX PUBLICS », et quelle force liante leur attribuer pour préciser le modèle, nous avons d'une part fait le choix de considérer que cette catégorie d'entité concerne des commerces de taille humaine. D'autre part, les informations concernant ces entités ayant été présélectionnées parmi l'immense quantité de données présente dans les actes d'enquête par l'analyste criminel lors de l'investigation, nous avons estimé que ces entités présentaient un lien avec l'affaire. Tenir compte de ces nouveaux éléments implique d'augmenter les forces liantes initialement attribuées (tableau 49).

Tableau 49 Correction des pondérations de symboles minimaux concernant des « LIEUX PUBLICS »

Symbole minimal	Pondération 1	Pondération 2
LIEU_PUBLIC-titulaire-TELECOM	0,1	0,4
PERSONNE-localisation-LIEU_PUBLIC	0,1	0,55
VEHICULE-localisation-LIEU_PUBLIC	0,08	0,4
PERSONNE-echange_objet_ou_argent-LIEU_PUBLIC	0,5	0,55

2. Pour les « PERSONNES MORALES »

Sont catégorisés comme des entreprises les bureaux, sièges sociaux et locaux. Nous avons ici considéré qu'il s'agit de grandes entreprises, telles qu'une chaîne de location de véhicule, dont les interactions avec le public ne sont pas directement liées à l'enquête, c'est-à-dire qu'elles se limitent aux services professionnels rendus par ces entreprises. Au regard de ce

choix, les forces liantes de certains symboles impliquant des personnes morales ont été légèrement diminuées (tableau 50).

Tableau 50 Correction des pondérations de symboles minimaux concernant des « PERSONNES MORALES »

Symbole minimal	Pondération 1	Pondération 2
PERSONNE_MORALE-titulaire-VÉHICULE	0,1	0,06
PERSONNE_MORALE-utilisateur-VÉHICULE	0,09	0,05

3. Pour les chaînes téléphoniques :

Enfin, une troisième anomalie nécessiterait un correctif dans le paramétrage : le traitement des liens téléphoniques. En effet, les symboles minimaux « PERSONNE-titulaire-TÉLÉCOM » et « TÉLÉCOM-en contact-TÉLÉCOM » ont tous deux des pondérations très élevées dans la mesure où les chaînes « PERSONNE – titulaire – TÉLÉCOM – en contact – TÉLÉCOM – titulaire – PERSONNE » soutiennent très fortement l’existence d’une connexion entre les deux individus. En revanche, dès lors qu’un *pattern* contient plusieurs fois le symbole « TÉLÉCOM-en contact-TÉLÉCOM »³⁷⁸, le chemin ne soutient plus l’existence du lien. La fonction d’agrégation gère la longueur des chemins : le fait de multiplier les forces liantes des symboles (comprises entre 0 et 1) fait rapidement décroître la pondération totale du chemin lorsque le nombre d’intermédiaires augmente. Le cas présenté ici est particulier : la décision va être d’accepter le lien pour un chemin de longueur 3, et de le rejeter pour les chemins de longueur supérieure à 3. Ainsi, dans le cas où le nombre de contacts téléphoniques intermédiaires augmente, la valeur de la pondération du chemin ne diminue pas suffisamment vite par rapport au changement de la décision sémantique dès la présence de deux contacts intermédiaires. Il s’agit là encore d’une limite de modèle, à laquelle nous n’avons pas pu remédier dans le cadre de la thèse.

4. Informations manquantes

L’analyse sémantique des chemins a affirmé la nécessité de disposer des attributs de dates et d’heures pour les relations, notamment pour les liens ponctuels³⁷⁹. En effet, la localisation de deux personnes au même endroit est considérablement moins informative sur

³⁷⁸ « PERSONNE – titulaire – TÉLÉCOM – en contact – TÉLÉCOM – en contact – TÉLÉCOM – titulaire – PERSONNE » (longueur 4).

³⁷⁹ Cf. section IV.1 b) « Détection graphique d’anomalies dans les données ».

une potentielle connexion s'il n'est pas possible de savoir si elles s'y trouvaient au même moment (figure 60).

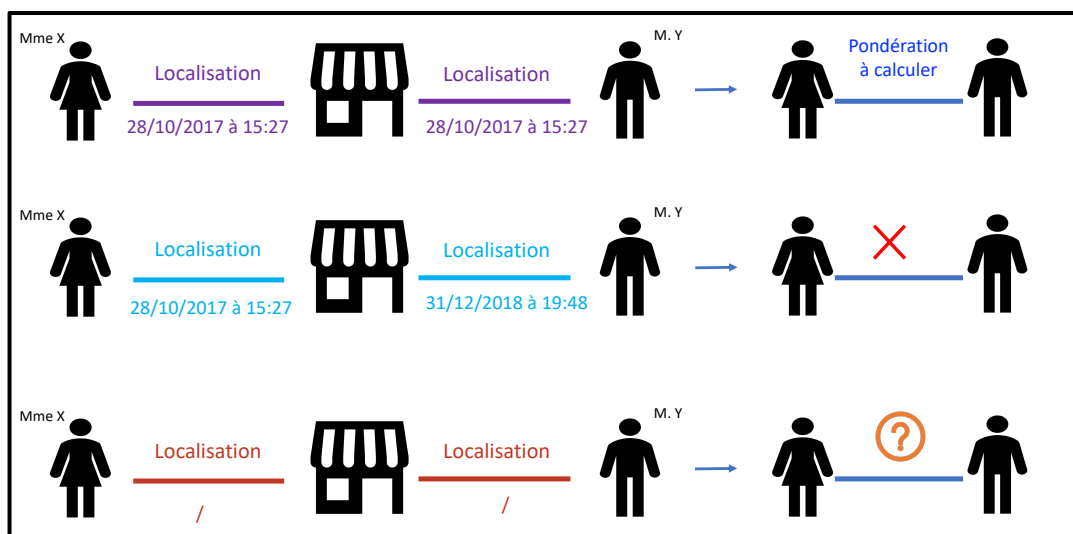


Figure 60 Nécessité de prendre en compte pour certains liens, tels que ceux de type « localisation », de la date et de l'heure

Or, ces deux informations essentielles sont absentes pour respectivement 59,02% et 65,90% des relations : ces taux sont trop importants pour pouvoir prendre en compte cette information de manière générique dans le modèle³⁸⁰. De fait, le modèle d'extraction du réseau social ne prend pas en compte ces données. Nous avons, lors de la conception de l'algorithme, conscience de ce biais non négligeable. Néanmoins, dans le temps imparti par les délais de thèse, nous avons décidé de nous concentrer sur l'élaboration de la méthodologie d'extraction, tout en sachant que des améliorations pourraient être ajoutées par la suite. Les réflexions concernant ces dernières sont développées dans la section VI.3 du présent chapitre.

À nouveau, pour les besoins de la recherche, une décision sur la gestion de ces liens a dû être prise en l'absence d'informations temporelles : rejeter systématiquement le lien ou calculer la pondération des chemins concernés « comme si » les dates et heures correspondaient. Nous avons choisi pour le développement de la méthode la deuxième option, en ayant conscience que cela induirait un biais non négligeable (mais difficilement quantifiable) sur les résultats³⁸¹. Le tableau 51 compare les pondérations des 25 chemins de l'échantillon avant et après corrections.

³⁸⁰ Le principe d'historisation mis en place grâce à *Tulip* permet, quand elles existent, de remonter à ces informations pour des liens d'intérêt.

³⁸¹ Un biais, également difficilement quantifiable, aurait aussi été induit dans le cas contraire.

Tableau 51 Décision sémantique et pondérations des 25 patterns de chemins étudiés avant (pondération1) et après (pondération2) corrections

Longueur	N° pattern	Décision	Pondération1	Pondération2
6	L6.5	Rejet	4,33E-04	1,69E-02
6	L6.4	Rejet	2,70E-05	2,43E-04
6	L6.3	Rejet	2,56E-04	1,21E-02
6	L6.2	Rejet	5,83E-03	5,83E-03
6	L6.1	Rejet	1,11E-03	3,46E-02
5	L5.5	Rejet	1,92E-04	2,88E-03
5	L5.4	Rejet	3,84E-05	2,64E-02
5	L5.3	Acceptation	1,05E-01	1,32E-01
5	L5.2	Rejet	2,34E-04	4,68E-05
5	L5.1	Rejet	5,07E-02	7,92E-02
4	L4.5	Acceptation	2,70E-03	6,76E-02
4	L4.4	Rejet	8,32E-02	1,20E-01
4	L4.3	Rejet	6,56E-01	6,56E-01
4	L4.2	Acceptation	3,24E-02	1,30E-01
4	L4.1	Rejet	4,23E-03	1,52E-03
3	L3.5	Acceptation	4,10E-01	3,78E-01
3	L3.4	Rejet	1,20E-02	1,20E-02
3	L3.3	Acceptation	5,20E-03	1,43E-01
3	L3.2	Rejet	2,60E-03	1,56E-02
3	L3.1	Acceptation	7,29E-01	7,29E-01
2	L2.5	Acceptation	4,23E-01	4,23E-01
2	L2.4	Rejet	1,60E-03	1,60E-03
2	L2.3	Acceptation	8,10E-01	8,10E-01
2	L2.2	Acceptation	6,40E-01	6,40E-01
2	L2.1	Acceptation	4,85E-01	4,85E-01

Dans la section suivante, les nouvelles valeurs de pondérations seront les valeurs après corrections. Nous nous intéresserons dans un premier temps aux performances du modèle d'extraction du réseau social après avoir effectué les correctifs. Dans un second temps, nous chercherons à déterminer la valeur seuil optimale pour filtrer les arêtes de *Social_network* et enfin obtenir le sociogramme de l'enquête, *Investigation_network*.

VI.2 Étude des performances et détermination du seuil de filtrage

La distribution des pondérations des 12 996 arêtes de *Social_network* n'est quasiment pas affectée par les corrections appliquées (figure 61).

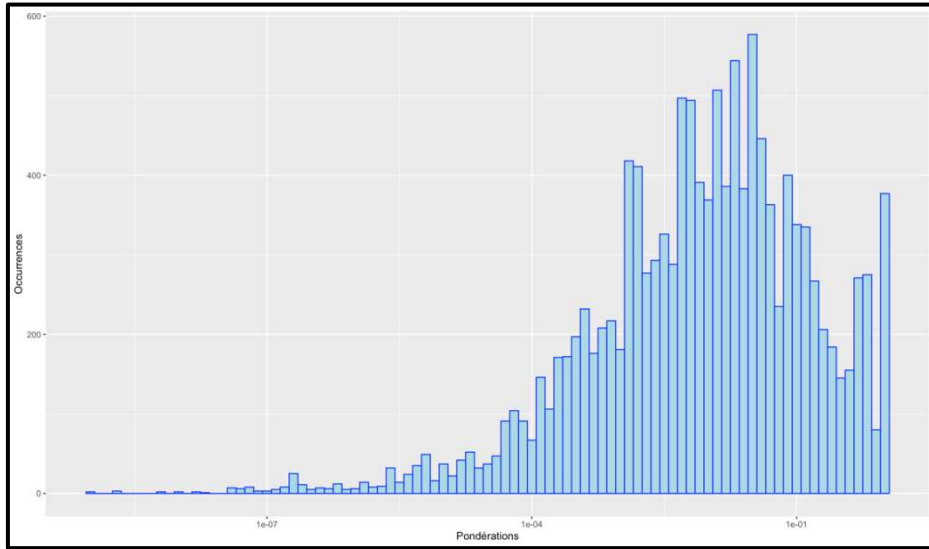


Figure 61 Distribution des pondérations des arêtes de *Social_network* (échelle x en \log_{10}) après corrections

a) Analyse des performances du modèle

Nous nous intéressons dans cette section aux distributions et aux densités respectives des pondérations des *patterns* de l'échantillon sémantiquement acceptés et rejetés, illustrées en figure 62.

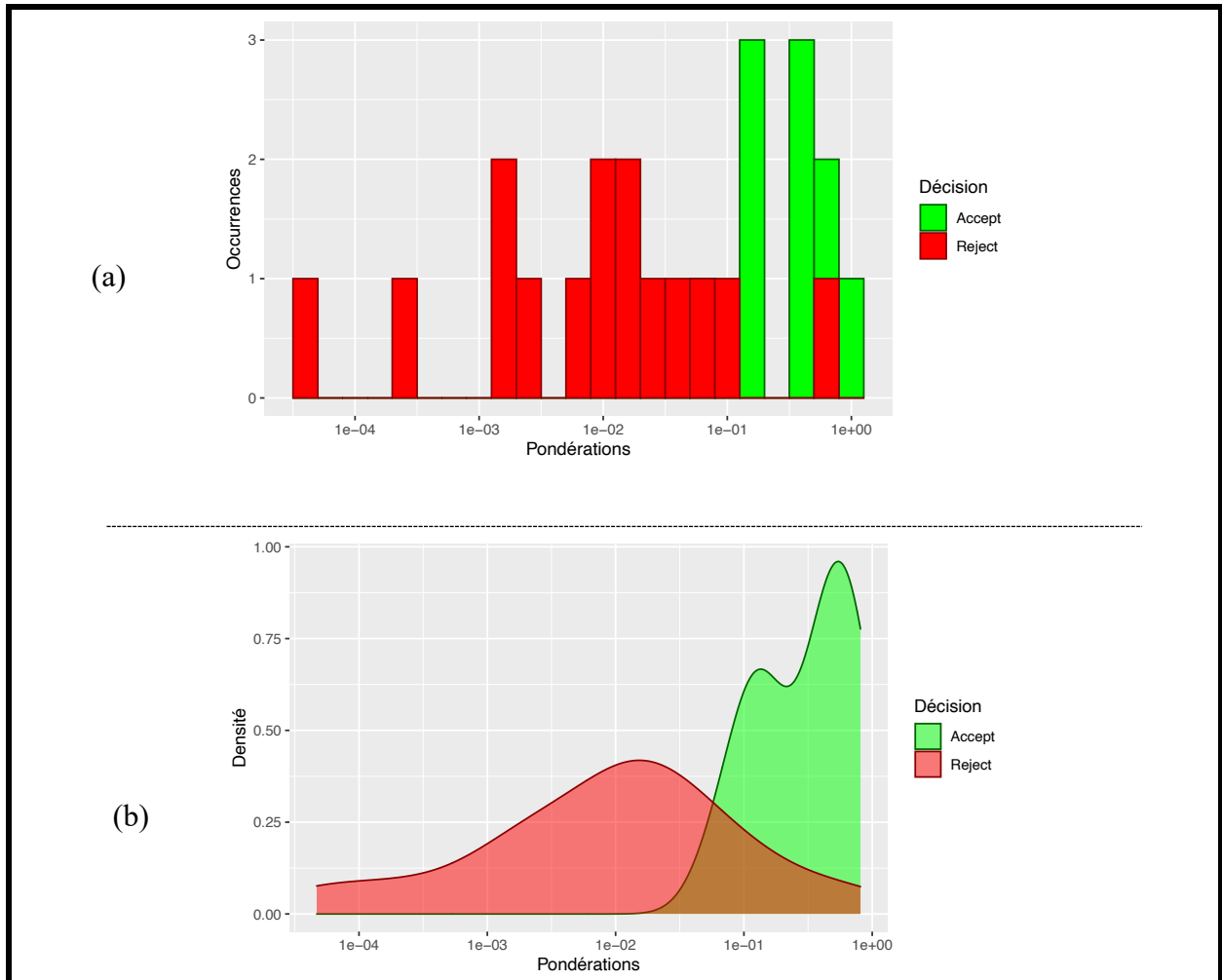


Figure 62 Distribution (a) et densité (b) des valeurs de pondération sur l'échantillon

Ces deux visualisations montrent qu'il existe un recouvrement entre les valeurs des pondérations des chemins sémantiquement acceptés et rejetés, c'est-à-dire que certaines valeurs correspondant à des *patterns* rejetés sont supérieures à certaines valeurs correspondant à des *patterns* acceptés (et inversement). Dès lors, quel que soit le seuil de filtrage fixé, la méthode présentera un taux d'erreur non nul. Il s'agit à présent de chercher la valeur seuil minimisant ces taux d'erreurs.

Les taux d'erreurs sont de deux natures : d'une part, les faux positifs, *i.e.* les *patterns* de chemins sémantiquement rejetés mais acceptés par le modèle, et d'autre part les faux négatifs, qui sont les *patterns* sémantiquement acceptés, mais rejetés par le modèle. Dans un premier temps, nous cherchons à savoir pour quelle valeur seuil le modèle va minimiser les erreurs de classification et maximiser les classifications correctes.

Pour comprendre, reprenons le tableau de comparaison des décisions sémantiques et des pondérations pour les 25 *patterns* de chemins de l'échantillon, et fixons un premier seuil arbitraire, $S_I = 0.1$ (tableau 52).

Tableau 52 Décision sémantique et pondération des 25 *patterns* de chemins étudiés

Longueur	N° pattern	Décision	Pondération
6	L6.5	Rejet	1,69E-02
6	L6.4	Rejet	2,43E-04
6	L6.3	Rejet	1,21E-02
6	L6.2	Rejet	5,83E-03
6	L6.1	Rejet	3,46E-02
5	L5.5	Rejet	2,88E-03
5	L5.4	Rejet	2,64E-02
5	L5.3	Acceptation	1,32E-01
5	L5.2	Rejet	4,68E-05
5	L5.1	Rejet	7,92E-02
4	L4.5	Acceptation	6,76E-02
4	L4.4	Rejet	1,20E-01
4	L4.3	Rejet	6,56E-01
4	L4.2	Acceptation	1,30E-01
4	L4.1	Rejet	1,52E-03
3	L3.5	Acceptation	3,78E-01
3	L3.4	Rejet	1,20E-02
3	L3.3	Acceptation	1,43E-01
3	L3.2	Rejet	1,56E-02
3	L3.1	Acceptation	7,29E-01
2	L2.5	Acceptation	4,23E-01
2	L2.4	Rejet	1,60E-03
2	L2.3	Acceptation	8,10E-01
2	L2.2	Acceptation	6,40E-01
2	L2.1	Acceptation	4,85E-01

Légende :
En rouge : valeurs rejetées (sémantiquement dans la colonne « Décision », par le modèle dans la colonne « Pondération »)
En vert : valeurs acceptées ((sémantiquement dans la colonne « Décision », par le modèle dans la colonne « Pondération »)
Fond jaune : erreurs pour $S_I = 0.1$

Les pondérations rejetées sont les valeurs strictement inférieures à S_I . Les pondérations acceptées sont celles dont les valeurs sont supérieures ou égales à S_I .

Nous voulons maintenant étudier d'une part la proportion de *patterns* sémantiquement acceptés qui sont effectivement acceptés et la proportion de *patterns* sémantiquement rejetés qui sont effectivement rejetés – *i.e.* les cas où le modèle affecte correctement les *patterns* à leur classe ; d'autre part la proportion de *patterns* sémantiquement rejetés et pourtant acceptés, et inversement, la proportion de *patterns* sémantiquement acceptés qui sont refusés par le modèle – *i.e.* les cas où le modèle affecte faussement les *patterns* à leur classe. Ces quatre indicateurs sont respectivement le taux de vrais positifs (TVP), le taux de vrais négatifs (TVN), le taux de faux positifs (TFP) et le taux de faux négatifs (TFN), où un *pattern* accepté est considéré comme positif, et un *pattern* rejeté comme négatif (tableau 53). Le tableau recensant ces quatre indicateurs est appelé « matrice de confusion » (tableaux 54 à 57).

Ils se calculent comme suit :

$$TVP = \frac{VP}{P} \quad \left| \quad TVN = \frac{VN}{N} \quad \left| \quad TFP = \frac{FP}{N} \quad \left| \quad TFN = \frac{FN}{P}$$

Tableau 53 **Légende**

VP	Nombre de vrais positifs
VN	Nombre de vrais négatifs
FP	Nombre de faux positifs
FN	Nombre de faux négatifs
P = VP + FN	Nombre de positifs sémantiques, qu'ils soient acceptés (VP) ou rejetés (FN) par le modèle
N = VN + FP	Nombre de négatifs sémantiques, qu'ils soient rejetés (VN) ou acceptés (FP) par le modèle

En se référant au tableau 52, nous avons, pour un seuil $S_I = 0.1$, les valeurs du tableau 54 dont on déduit la matrice de confusion (tableau 55).

Tableau 54 Décompte des VP, VN, FP, FN pour S_1

VP	9
VN	13
FP	2
FN	1
P = VP + FN	10
N = VN + FP	15

Tableau 55 Matrice de confusion pour $S_1 = 0.1$

		Décision du modèle	
		Acceptation	Rejet
Décision sémantique (Classe)	Acceptation	TVP = 90%	TFN = 10%
	Rejet	TFP = 13,3%	TVN = 86,6%

La sensibilité, qui correspond au TVP, est ici de 90%, ce qui signifie qu'un *pattern* accepté sémantiquement pris au hasard parmi l'échantillon a une probabilité de 90% d'être accepté par le modèle et de 10% d'être refusé (ces deux grandeurs étant corrélées : $TVP = 1 - TFN$). Le taux de vrais négatifs, également appelé spécificité, est de 86,6%, ce qui signifie qu'un *pattern* sémantiquement rejeté pris au hasard parmi l'échantillon a une probabilité de 86,6% d'être rejeté par le modèle, et de 13,3% d'être accepté ($TVN = 1 - TFP$).

Fixons à présent un deuxième seuil arbitraire : $S_2 = 0.06$. Nous obtenons alors les résultats des tableaux 56 et 57.

Tableau 56 Décompte des VP, VN, FP, FN pour S_2

VP	10
VN	12
FP	3
FN	0
P = VP + FN	10
N = VN + FP	15

Tableau 57 Matrice de confusion pour $S_2 = 0.06$

		Décision du modèle	
		Acceptation	Rejet
Décision sémantique (Classe)	Acceptation	TVP = 100%	TFN = 0%
	Rejet	TFP = 20%	TVN = 80%

La sensibilité est passée de 90% à 100%. En ne prenant en compte que cet indicateur, il serait tentant de conclure directement que le modèle est plus performant avec S_2 qu'avec S_1 .

Néanmoins, la spécificité est passée de 86,6% à 80%. Ainsi, le modèle avec S_2 est plus sensible, tandis que le modèle avec le seuil S_1 est plus spécifique. Dès lors, nous allons chercher à visualiser l'évolution de ces deux grandeurs pour différentes valeurs seuil. Cela peut être fait en traçant une courbe ROC (*Receiver Operating Characteristics*), qui représente, pour différentes valeurs seuils, la sensibilité en fonction de (1 - spécificité), c'est-à-dire le taux de vrais positifs en fonction du taux de faux positifs, comme illustré en figure 63. Les deux axes (x, y) de la figure sont bornés entre 0 et 1³⁸². Les calculs ont été effectués à l'aide de l'interface *RStudio*³⁸³, en langage *R*. Plus précisément, les fonctions `prediction` et `performance` de la librairie (`ROCR`) ont été utilisées. La première permet de formater les données d'entrée (dans notre cas, les colonnes « Décision » et « Pondération » du tableau 1 en un objet *R* de la classe « `prediction` » permettant de contenir les prédictions numériques avec les étiquettes associées de classe réelles (Acceptation / Rejet). Cette fonction génère également les seuils à évaluer. Ce type d'objet est ensuite géré par la deuxième fonction, `performance`, qui permet d'effectuer de nombreuses évaluations de prédicteurs et de tracer la courbe ROC. Cette dernière est définie par un nombre de point égal au nombre de valeurs seuils calculées³⁸⁴ (figure 63).

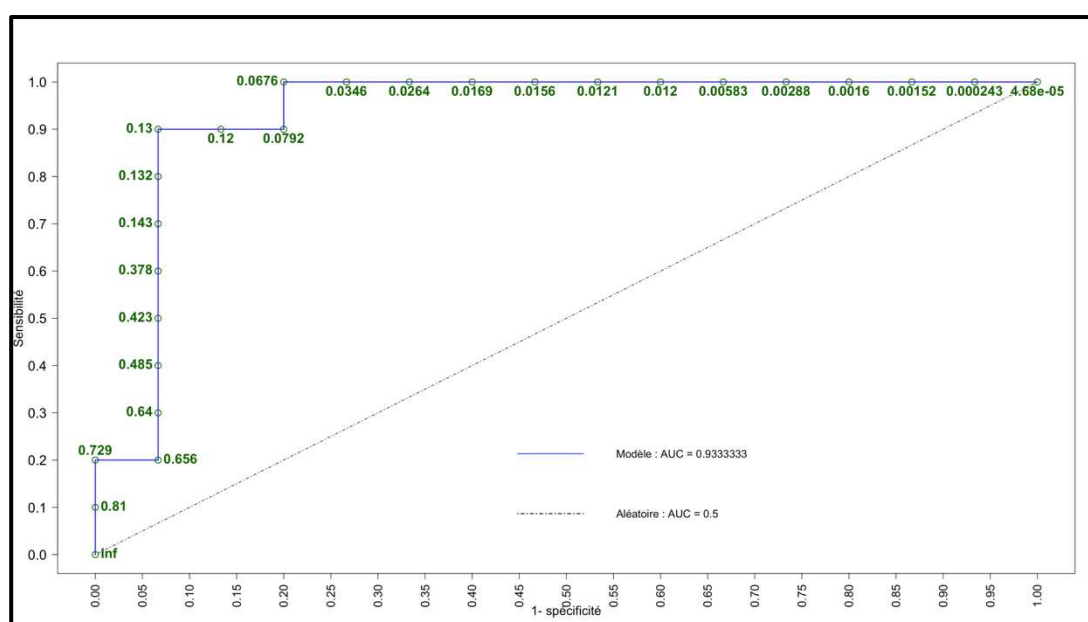


Figure 63 Courbe ROC illustrant le TVP en fonction du TFP pour les différentes valeurs seuils (en vert)

La figure 62 a montré l'existence d'un recouvrement entre les distributions des *patterns* de chemins sémantiquement acceptés et rejetés. L'aire sous la courbe ROC (*Area Under Curve* – AUC) permet d'évaluer la zone de chevauchement : c'est une mesure directe de la séparation

³⁸² Par définition, $0 \leq \text{TVP} \leq 1$ et $0 \leq \text{TFN} \leq 1$.

³⁸³ Cf. section III.4 c) « Traitements de données : *Excel* et *RStudio* ».

³⁸⁴ Tobias Sing, Oliver Sander, Niko Beerenwinkel, Thomas Lengauer, Thomas Unterthiner et Felix G. M. Ernst, « ROC: Visualizing the Performance of Scoring Classifiers ».

entre les deux classes³⁸⁵. Une AUC égale à 1 correspond à une séparation parfaite entre les deux classes, tandis qu'une AUC égale à 0 indique une séparation parfaite mais inversée. Une AUC à 0.5 (représentée par la diagonale de la figure 63) se rapporte à une classification aléatoire des *patterns* dans l'une ou l'autre des classes. Finalement, l'AUC équivaut à la probabilité qu'une valeur prise au hasard dans l'échantillon soit correctement classifiée par le modèle (VP ou VN) : plus la valeur est proche de 1, plus le modèle est capable de distinguer les *patterns* sémantiquement acceptés et rejetés³⁸⁶.

b) Détermination du seuil de filtrage des pondérations

Les performances du modèle étant estimées, un seuil de filtrage des arêtes doit être déterminé afin de finaliser l'extraction du réseau social de l'enquête.

Ont été évoqués dans la section précédente quatre indicateurs permettant d'évaluer le succès d'un modèle pour un seuil donné : la sensibilité (ou taux de vrais positifs), la spécificité (ou taux de vrais négatifs), les taux de faux positifs et de faux négatifs ainsi que leurs relations. Il en existe de nombreux autres, tels que la précision (*prec*), le taux d'erreur global (*err*), les valeurs de prédiction positive (VPP) et négative (VPN)³⁸⁷, les rapports de vraisemblance positif (RVP) et négatif (RVN)³⁸⁸. Un tableau détaillé de ces indicateurs et de leurs interprétations est disponible en annexe 13.

$$prec = \frac{VP+VN}{P+N} \qquad err = \frac{FP+FN}{P+N}$$

En fonction du cas d'application, du besoin, des objectifs, et des coûts de classifications erronées, la chercheuse ou l'analyste va rechercher la valeur seuil qui maximise ou minimise les uns ou les autres de ces indicateurs. Ainsi, le choix du seuil de décision découle de la question traitée. En épidémiologie, ce sera typiquement la valeur qui minimise le taux de faux

³⁸⁵ Miguel de Figueiredo, *Évaluation d'une approche chimiométrique non ciblée pour l'inférence de source de liquides inflammables en science forensique*, Université Paris-Saclay, Paris, 2018, 259 p.

³⁸⁶ Pierre Esseiva, Laeticia Gaste, Daniel Alvarez et Frederic Anglada, « Illicit drug profiling, reflection on statistical comparisons », *Forensic Science International*, vol. 207, n° 1, 2011, pp. 27-34.

³⁸⁷ Tobias Sing, Oliver Sander, Niko Beerenwinkel, Thomas Lengauer, Thomas Unterthiner et Felix G. M. Ernst, « ROCR », *op. cit.*

³⁸⁸ Daisy Bertrand, Joel Fluss, Catherine Billard et Johannes C. Ziegler, « Efficacité, sensibilité, spécificité : comparaison de différents tests de lecture », *L'Année psychologique*, vol. 110, n° 2, 2010, pp. 299-320.

négatifs, au risque d’avoir un taux de faux positifs légèrement élevé : la détection d’un patient malade comme sain aura des conséquences plus graves que la détection d’un patient sain comme malade. En criminalistique chimique ou en sciences forensiques, c’est plutôt l’équilibre entre un taux de faux positifs minimal et un taux de vrais positifs maximal qui va être recherché³⁸⁹.

En s’appuyant sur les connaissances acquises lors du stage « Analyste en renseignement criminel » effectué au SCRC en septembre 2019 et sur l’expérience du Lieutenant-Colonel Jérôme Barlatier dont la doctorante a bénéficié tout au long de sa thèse au cours de nombreux et fructueux échanges, il est apparu que les critères de sélection du seuil doivent dépendre du besoin de l’analyste, des problématiques, des enjeux et des coûts auxquels elle est confrontée. En effet, selon Maurice Cusson, le renseignement criminel peut être utilisé d’au moins huit manières³⁹⁰, dont deux nous intéressent spécifiquement : le soutien et l’orientation de l’enquête judiciaire et l’analyse de phénomènes criminels dans un but proactif (tableau 58).

Tableau 58 Problématiques distinctes dans des logiques d’enquête et de renseignement³⁹¹

Soutien à l’enquête	Analyse de phénomènes
Réactif, au cas par cas	Proactif, global
Aide à la reconstruction des événements pour fonder une vérité judiciaire à partir d’éléments collectés en procédure (conformément au droit) et destiné à constituer une preuve	Lutter contre l’incertitude pour générer de la connaissance à partir d’éléments d’information
La déclaration des faits entraîne l’action pour faire cesser l’évènement	Le savoir guide l’action sur les phénomènes criminels
Réponses répressives, répression pénale	Réponses proactives, régulation sociale

Ainsi, dans une logique de renseignement en soutien et en orientation de l’enquête, l’analyste criminel veut éviter la déperdition d’éléments : il va privilégier un faible taux de faux négatif. Non pas que le taux de faux positifs ne puisse être élevé sans conséquences, mais dans la mesure où le modèle conservera seulement environ 10% des données, il est plus aisé d’effectuer un contrôle manuel des liens ambigus sur les chemins acceptés pour détecter des

³⁸⁹ Pierre Esseiva, Laeticia Gaste, Daniel Alvarez et Frederic Anglada, « Illicit drug profiling, reflection on statistical comparisons », *op. cit.*

³⁹⁰ Maurice Cusson, Frédéric Lemieux et Benoît Dupont, *Traité de sécurité intérieure, op. cit.*

³⁹¹ Jérôme Barlatier, « Criminal Investigation and Criminal Intelligence: Example of Adaptation in the Prevention and Repression of Cybercrime », *Risks*, vol. 8, n° 3, 18/09/2020, pp. 99-109.

faux positifs que d'analyser les 90% de liens rejetés à la recherche de faux négatifs. En outre, dans une perspective d'appui à l'investigation, où une seule affaire est traitée, le volume de liens ambigus autorise à envisager ce contrôle manuel lorsque cela est nécessaire. Enfin, le réseau social extrait est un modèle des connexions existantes entre les personnes impliquées dans l'affaire, ce n'est pas un élément probant au sens de l'enquête de l'enquête judiciaire. L'analyste criminel « n'arrête pas, n'accuse pas » : il oriente l'enquête sur de potentielles connexions, et vers les individus susceptibles de détenir une fonction ou une position stratégique pour l'organisation criminelle. La suite de l'enquête pourra révéler que la connexion n'existe finalement pas, ou n'a que peu d'importance. Ainsi, le coût d'un faux négatif est une perte d'information pourtant à disposition permettant potentiellement d'aboutir à des éléments probants, tandis que le coût d'un faux positif se verra décroître au fur-et-à-mesure de l'enquête, lorsque les investigations infirmeront l'existence d'une relation établie à tort.

Dans une logique de renseignement criminel pour l'analyse de phénomènes, l'analyste est souvent amené à traiter plusieurs affaires afin de détecter des tendances dans les schémas sociaux, fonctionnels et organisationnels récurrents, permettant ainsi de capitaliser des connaissances sur les mécanismes sociaux à l'œuvre dans un champ criminel donné. Dans de tels cas, le volume de données à disposition rend inenvisageable les contrôles manuels : l'analyste va donc chercher à minimiser le taux d'erreur (*err*) global et maximiser la précision (*prec*) du modèle sur l'ensemble des données afin d'en extraire le plus d'information possible. Dans cette optique, il ne s'agit pas d'identifier des individus³⁹² mais de détecter et d'analyser des *patterns* d'interactions sociales au sein de réseaux criminels : l'analyse n'est pas suivie d'une investigation criminelle confirmant ou invalidant les résultats. En ce sens, le renseignement consiste à capitaliser des connaissances sur des phénomènes passés afin de comprendre des mécanismes criminels, d'adapter les réponses proactives et d'orienter les enquêtes futures. Ainsi, l'analyste va rechercher l'équilibre optimal entre la sensibilité et la spécificité.

Le tableau 59 illustre, pour chacun des seuils évalués par la fonction *prediction*, la sensibilité, la spécificité, les taux de faux positifs, faux négatifs et d'erreur globale ainsi que la précision.

³⁹² Typiquement, le présent travail de renseignement criminel est effectué sur des données complètement anonymisées.

Tableau 59 Calcul de différents indicateurs pour chaque valeur seuil évaluée

Seuils	Sensibilité	Spécificité	TFP	TFN	Erreur globale	Précision
Inf	0,00	1,00	0,00%	100,00%	40,00%	0,00
8,100E-01	0,10	1,00	0,00%	90,00%	36,00%	0,10
7,290E-01	0,20	1,00	0,00%	80,00%	32,00%	0,20
6,560E-01	0,20	0,93	6,67%	80,00%	36,00%	0,20
6,400E-01	0,30	0,93	6,67%	70,00%	32,00%	0,30
4,850E-01	0,40	0,93	6,67%	60,00%	28,00%	0,40
4,230E-01	0,50	0,93	6,67%	50,00%	24,00%	0,50
3,780E-01	0,60	0,93	6,67%	40,00%	20,00%	0,60
1,430E-01	0,70	0,93	6,67%	30,00%	16,00%	0,70
1,320E-01	0,80	0,93	6,67%	20,00%	12,00%	0,80
1,300E-01	0,90	0,93	6,67%	10,00%	8,00%	90,00%
1,200E-01	0,90	0,87	13,33%	10,00%	12,00%	0,90
7,920E-02	0,90	0,80	20,00%	10,00%	16,00%	0,90
6,760E-02	1,00	0,80	20,00%	0,00%	12,00%	1,00
3,460E-02	1,00	0,73	26,67%	0,00%	16,00%	1,00
2,640E-02	1,00	0,67	33,33%	0,00%	20,00%	1,00
1,690E-02	1,00	0,60	40,00%	0,00%	24,00%	1,00
1,560E-02	1,00	0,53	46,67%	0,00%	28,00%	1,00
1,210E-02	1,00	0,47	53,33%	0,00%	32,00%	1,00
1,200E-02	1,00	0,40	60,00%	0,00%	36,00%	1,00
5,830E-03	1,00	0,33	66,67%	0,00%	40,00%	1,00
2,880E-03	1,00	0,27	73,33%	0,00%	44,00%	1,00
1,600E-03	1,00	0,20	80,00%	0,00%	48,00%	1,00
1,520E-03	1,00	0,13	86,67%	0,00%	52,00%	1,00
2,430E-04	1,00	0,07	93,33%	0,00%	56,00%	1,00
4,680E-05	1,00	0,00	100,00%	0,00%	60,00%	1,00

Légende :
 En bleu : Valeurs optimales dans une démarche d'orientation de l'enquête criminelle
 En violet : Valeurs optimales dans une démarche de détection de tendance

Dans une logique d'enquête, l'analyste sélectionnera le seuil qui minimisera le taux de faux négatifs, tout en maintenant un taux de faux positif raisonnable. La méthode dite « du coin supérieur gauche » (CSG)³⁹³ est souvent utilisée pour ce faire. Elle permet d'identifier le seuil pour lequel le point de la courbe ROC est le plus près du coin supérieur gauche du graphique, ayant pour coordonnées (0 ;1)³⁹⁴. L'indice du CSG se calcule ainsi :

$$CSG = \max[(1 - \text{sensibilité})^2 + (1 - \text{spécificité})^2]$$

Il s'agit du point $S_0 = 0.13$, correspondant à un taux de faux négatifs de 10% et à un taux de vrais positifs de 90%. Dans notre cas particulier, après évaluation avec le Ltd-Col J. Barlatier et pour les raisons énoncées plus haut, nous allons cependant choisir le premier seuil correspondant à un TFN inférieur à celui de S_0 : le point $S_{enq} = 0.676$, dont le TFN est de 0% et le TFP de 20% (TVP = 100%, TVN = 80%). En filtrant le graphe *Social_network* avec le seuil

³⁹³ Miguel de Figueiredo, *Évaluation d'une approche chimiométrique non ciblée pour l'inférence de source de liquides inflammables en science forensique*, op. cit.

³⁹⁴ L'abscisse représente le taux de faux négatifs et l'ordonnée le taux de vrais positifs. Le point (0, 1) correspond donc à un modèle idéal où il existe un seuil pour lequel le taux de faux négatifs est nul et la sensibilité vaut 1.

S_{enq} , le graphe *Investigation_network_{enq}* est obtenu (figure 64). Il contient 665 personnes pour 3 096 arêtes dont la pondération est supérieure à S_{enq} .

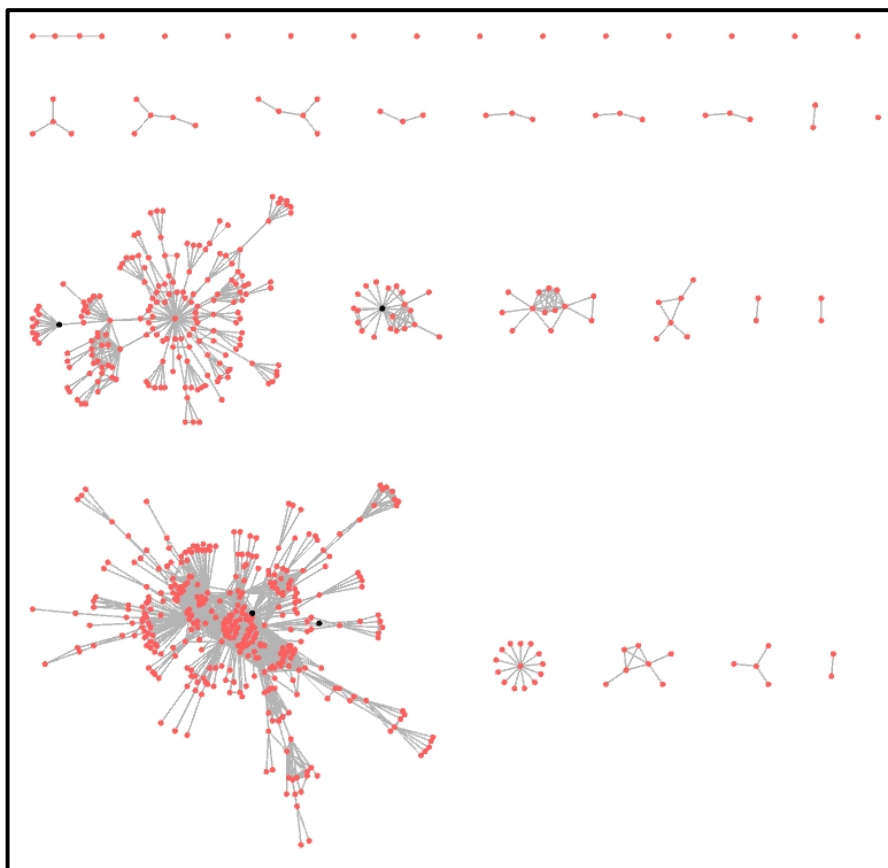


Figure 64 Graphe *Investigation_network_{enq}*; N = 665, L=3 096

Dans une logique d'analyse de phénomènes criminels, l'analyste sélectionnera plutôt le seuil minimisant le taux d'erreur global et maximisant la précision. Il s'agit du point $S_{rens} = 0.13$, correspondant à une erreur globale de 8% et une précision de 90%, pour un TVP de 90% et un TVN 93% (TFP = 6.67%, TFN = 10%). En filtrant *Social_network* avec le seuil S_{rens} , le graphe *Investigation_network_{rens}* est obtenu (figure 65). Il contient 665 personnes pour 2 086 arêtes dont la pondération est supérieure à S_{rens} .

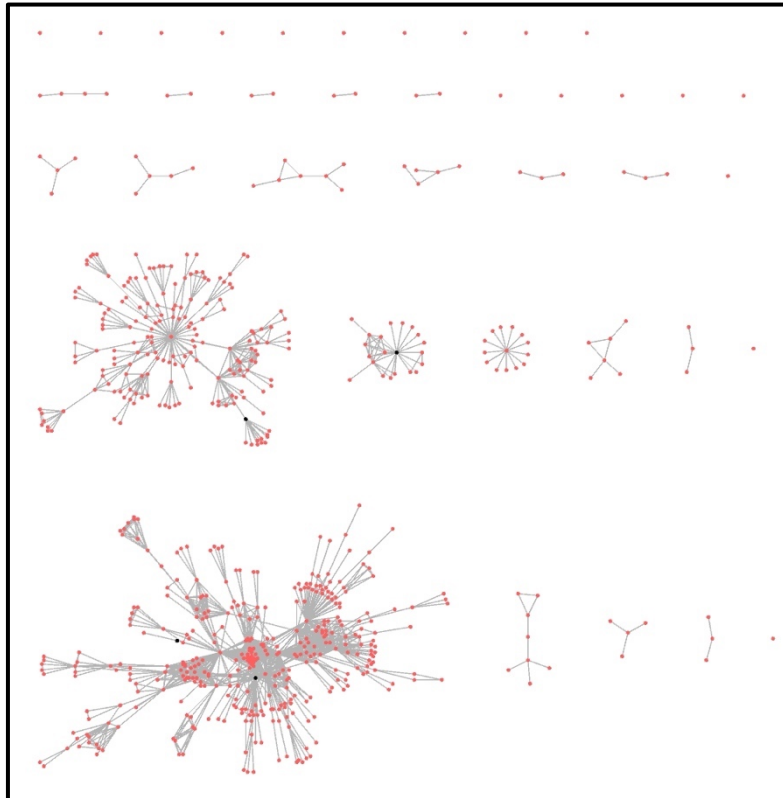


Figure 65 Graphe *Investigation_networkrens*; $N = 665$, $L = 2\ 086$

La figure 66 représente la position des seuils S_{enq} et S_{rens} sur la courbe ROC du modèle.

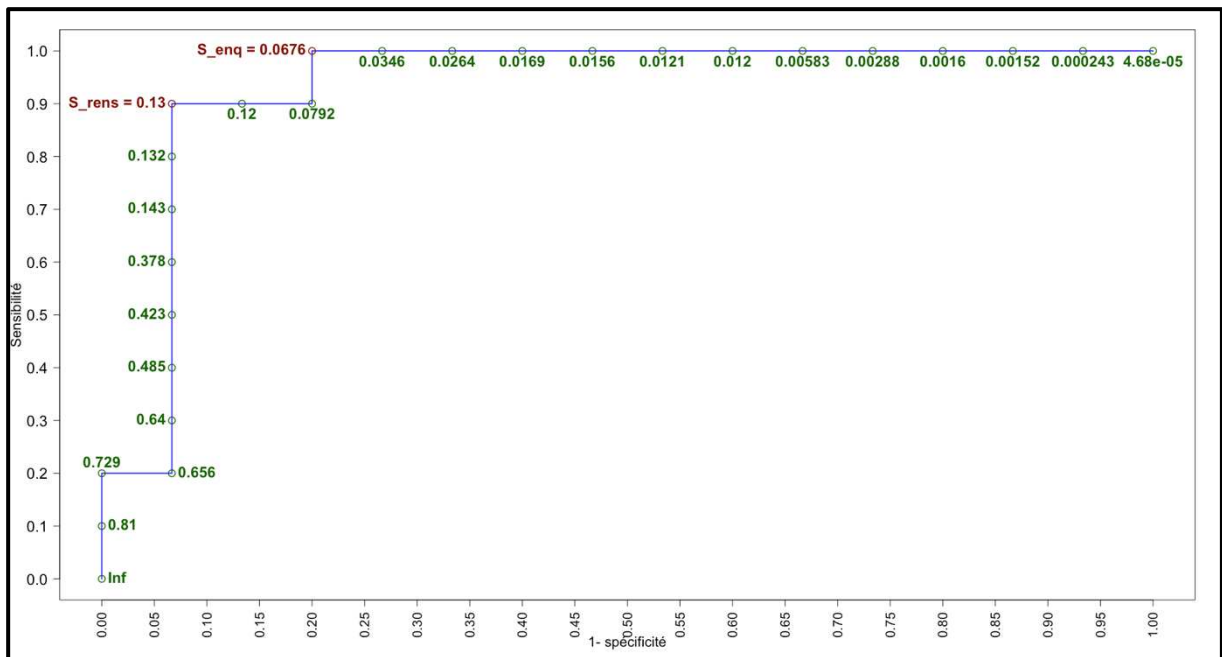


Figure 66 Position des seuils sur la courbe ROC

Les analyses de réseaux sociaux seront effectuées dans le chapitre suivant, puis interprétées et discutées dans le septième chapitre.

VI.3 Discussion et élaboration de critères d'optimisation

Les résultats de l'étude de la performance du modèle sont encourageants. Néanmoins, la méthode possède des limites, partiellement évoquées dans la section VI.1. La suite de ce chapitre s'attache à les rappeler et propose des éléments de solution.

a) Limites du modèle

La première limite que nous avons évoquée est le nombre de *patterns* considérés dans l'échantillon de test des performances du modèle. En effet, les tests ont été effectués sur 25 *patterns* de chemins, ce qui est très faible et statistiquement très peu représentatif de la population globale. Il existe plus de 7 253 *patterns*³⁹⁵ de chemins induisant des liens directs entre individus. Néanmoins, seulement 276 *patterns* de longueur inférieure ou égale à 6 – longueur maximale de l'échantillon – ont été détectés avant d'arriver à saturation des serveurs. L'échantillon représente donc 9,06% des *patterns* connus. Idéalement, il aurait fallu pouvoir étudier un échantillon d'au moins 100 motifs. Cependant, l'étude des occurrences des symboles minimaux présents dans les 25 *patterns* dans le fichier initial de données montre que si l'échantillon n'est pas représentatif des données en termes de motifs de chemins, c'est-à-dire en termes de combinaison de symboles minimaux sur les chemins entiers, il est représentatif des symboles minimaux existant dans le graphe *data0*. En effet, sur les 69 symboles initiaux, 12 concernent des liens directement établis entre des personnes lors de l'investigation, et peuvent être écartés de la validation du modèle³⁹⁶. Ainsi, les *patterns* à étudier sont des combinaisons des 57 symboles minimaux restants. Les 13 symboles les plus fréquents sont présents dans l'échantillon (*cf.* annexes 5, 10 et 11). Les 20 symboles minimaux de l'échantillon³⁹⁷ ont une occurrence cumulée de 1 638 dans *data0* sur 2 357 liens initiaux, ils représentent donc 69,5% de la population de symboles minimaux. Les fonctions de fusion et d'agrégation étant commutatives, il est possible, bien que nous n'ayons testé que 25

³⁹⁵ Notre première tentative d'extraction étant tombée en erreur après avoir extrait 7 253 *patterns* de chemins, nous ignorons leur nombre exact.

³⁹⁶ Ces symboles sont pour 9 d'entre eux pondérés à 1 (PERSONNE-familial/amical-PERSONNE). Ils ne sont pas considérés dans des chemins, puisqu'ils représentent déjà un lien direct entre individus. Les 3 restants sont pondérés à 0 (PERSONNE-non lien-PERSONNE), ils n'induisent donc pas de lien direct entre individus dans *Social_network*.

³⁹⁷ *Cf.* annexe 10 « Détail des *patterns* de chemins évalués (échantillon) pour l'analyse des résultats de l'extraction du réseau social et la détermination de la valeur seuil de filtrage du réseau social d'enquête ».

combinaisons, que l'échantillon soit finalement plus représentatif de la population totale de *patterns* qu'il ne semble au premier abord.

Une deuxième limite concerne la précision des catégories établies dans les ontologies. Nous avons montré dans la section VI.1 que certaines catégories, telles que « LIEU PUBLIC » et « PERSONNE MORALE » s'avéraient problématiques lors de l'analyse sémantique des chemins dans la mesure où elles n'étaient pas assez restrictives. Ces deux catégories font référence à des entités en lien avec une activité professionnelle, mais sont considérées distinctement. La première concerne les lieux accueillant du public, tels que des commerces et restaurants, la seconde concerne les bureaux et locaux. Ne disposant pas d'informations sur la taille des entreprises ni sur leur potentielle implication dans l'affaire, nous avons ainsi dû fixer ces critères, et admettons dans la suite du travail que les « LIEUX PUBLICS » sont de taille humaine et peuvent avoir un lien direct avec l'enquête, tandis que les « PERSONNES MORALES » sont de grandes entreprises dont les interactions avec le public ne sont pas directement liées à l'enquête. Ainsi, créer de nouvelles catégories plus restrictives et prenant en compte ces informations permettrait de diminuer l'incertitude et ainsi les taux d'erreurs. Cependant, le nombre de catégories doit rester suffisamment faible pour que l'analyste puisse attribuer les forces liantes aux symboles minimaux résultants en un temps raisonnable. En outre, si ces distinctions peuvent apporter de la précision au modèle, elles peuvent aussi en biaiser les résultats dans la mesure où certaines entités pourraient être difficilement classifiables dans l'une ou l'autre de ces catégories restrictives. Notamment, un commerce donné pourrait être considéré pour certains types liens comme un lieu (lorsque le lien est de type « localisation », ou « échange objet ou argent ») et pour d'autres comme une personne morale (tels que les liens de type « professionnel » ou « titulaire »).

Là encore, l'approche peut être différenciée en fonction de la logique de renseignement criminel dans laquelle l'analyste se trouve. Dans une démarche de soutien à l'enquête, le modèle doit être facilement applicable, efficace et efficient : l'analyste ne peut pas, à chaque affaire, passer un temps considérable sur la catégorisation des liens et entités ni sur la pondération des symboles minimaux résultants. Dans une démarche proactive d'analyse de phénomènes, l'analyste traite simultanément plusieurs affaires similaires et travaille sur des temporalités différentes. Il peut donc effectuer une analyse plus approfondie en appliquant une catégorisation certes plus chronophage mais plus précise, et bénéficier ainsi d'un modèle plus performant.

Nous avons également évoqué dans la section VI.1 la problématique des chaînes téléphoniques. Étant constituées de successions de liens de forces liantes très élevées (égales à 0.9), ces chaînes induisent dans *Social_network* des arêtes avec une pondération importante et

donc des liens directs acceptés par le modèle, alors même que ces chaînes sont sémantiquement rejetées à partir d'une longueur de 4³⁹⁸.

La quatrième limite concerne l'absence d'informations de date et d'heure sur respectivement 59,02% et 65,90% des données relationnelles originales. Or, dans beaucoup de cas, l'établissement de liens directs ne dépend pas uniquement de la sémantique des chemins dont ils sont issus, mais aussi de la proximité temporelle – ou non – des faits (figure 60).

Ces limites sont notamment problématiques car l'algorithme ne permet pas actuellement d'extraire de manière efficiente dans un format facilement exploitable les chemins du graphe *data0* ayant induit les liens directs de *Social_Network*. En effet, le modèle est doté d'un script permettant à l'analyste de supprimer manuellement des liens directs conservés qu'il estime ne pas exister sur le terrain. Ainsi, pouvoir remonter aux données originales pour les liens directs sur lesquels l'analyste a un doute lui permettrait de contrôler ces derniers, et de supprimer du sociogramme *Investigation_network_{enq}* (ou *Investigation_network_{rens}*) les liens qu'il décide finalement de rejeter.

Enfin, le modèle n'est pour l'instant pas généralisable à d'autres cas d'usage, dans la mesure où les forces liantes n'ont été attribuées qu'aux symboles minimaux existants dans les données. Pour l'appliquer par exemple sur une enquête portant sur du trafic d'armes à feu, il faudrait effectuer une nouvelle étape de pondération des nouveaux symboles, tels que {PERSONNE – acheteur – ARME À FEU}. Néanmoins, cela reste envisageable dans une démarche de capitalisation de connaissances, dans laquelle, à chaque nouveau cas d'usage, la base de données de symboles minimaux serait alimentée et augmentée jusqu'à couvrir un large spectre de cas d'applications.

Les limites énoncées influent sur la validité du modèle. Solutionner ces points d'attentions permettrait de réduire la zone de chevauchement des pondérations des liens directs sémantiquement acceptés et rejetés – et donc d'améliorer la capacité du modèle à faire la séparation entre les deux classes.

³⁹⁸ Les liens directs induits par les chemins {PERSONNE – titulaire/utilisateur – TÉLÉCOM – en contact – TÉLÉCOM – en contact – TÉLÉCOM – titulaire/utilisateur – PERSONNE} de longueur 4 ou supérieure sont sémantiquement rejetés.

b) Éléments de solutions : optimisation de l'algorithme

Certaines des limites soulevées pourraient être résolues par des développements futurs de l'algorithme.

Un premier point d'amélioration concernerait les entités qui, en fonction des types de liens auxquels elles sont associées, pourraient être classifiées tantôt dans une catégorie, tantôt dans une autre. Une solution pourrait être de créer deux nœuds pour la même entité et de les réunir dans un méta-nœud, de la même manière que pour les personnes ayant des identités multiples³⁹⁹. La démarche suggérée est illustrée en figures 67 et 68.

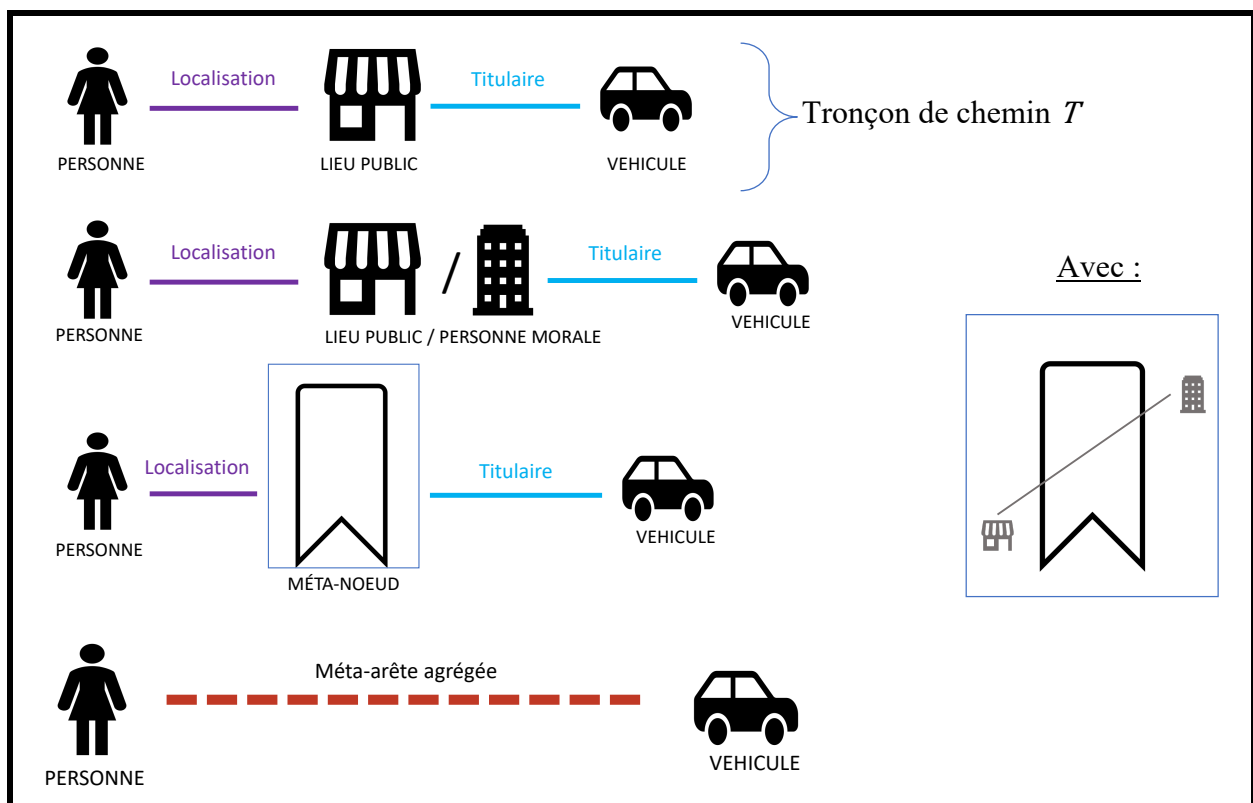


Figure 67 Stockage des entités difficilement classifiables dans des méta-nœuds

L'algorithme pourrait alors traiter le tronçon de chemin T et calculer la pondération de la méta-arête comme indiqué dans la figure 68.

³⁹⁹ Cf. section IV.1 a) « Création du graphe relationnel n -mode », figure 22.

$$T: \{ \text{PERSONNE} - \text{méta-arête agrégée} - \text{VÉHICULE} \}$$

$$T: \{ \text{PERSONNE} - \text{localisation} - \{ \text{LIEU PUBLIC} / \text{PERSONNE MORALE} \} - \text{titulaire} - \text{VÉHICULE} \}$$

$$T: \{ \text{PERSONNE} - \text{localisation} - \text{LIEU PUBLIC} \} - \{ \text{PERSONNE MORALE} - \text{titulaire} - \text{VÉHICULE} \}$$

Avec :

$$p_1 = p(\text{PERSONNE} - \text{localisation} - \text{LIEU PUBLIC})$$

$$p_2 = p(\text{PERSONNE MORALE} - \text{titulaire} - \text{VÉHICULE})$$

$$P_T = P_{\text{méta-arête}} = p_1 \cdot p_2$$

Figure 68 Calcul de la pondération du tronçon de chemin T

Une solution peut également être imaginée pour la gestion des chaînes téléphoniques de longueur supérieure à 4. Cette dernière consisterait à introduire une condition supplémentaire dans l'algorithme lors de l'agrégation des trinômes de ces chemins en liens directs ; ainsi qu'un nouvel attribut, nommé « alerte », aux liens directs, indiquant s'ils sont issus de chaînes devant faire l'objet d'un contrôle par l'analyste. Dans les cas où un lien direct est induit par un *pattern* de type {PERSONNE – utilisateur – TÉLÉCOM- en contact – TÉLÉCOM – en contact – TÉLÉCOM – titulaire – PERSONNE}, l'attribut « alerte » du lien direct prendrait la valeur 1, et la valeur 0 sinon. Le calcul de la valeur de « alerte » serait alors effectué en posant comme condition que si les forces liantes de l'ensemble des symboles composant le lien direct sont supérieures ou égales à 0.85, alors l'attribut « alerte » prend la valeur 1, et prend la valeur 0 sinon. L'analyste pourrait alors contrôler manuellement les liens directs dont la valeur de « alerte » vaut 1 et les supprimer du graphe si besoin (figure 69).

$$L: \{ \text{PERSONNE} - \text{utilisateur} - \text{TÉLÉCOM} - \text{en contact} - \text{TÉLÉCOM} - \text{en contact} - \text{TÉLÉCOM} - \text{titulaire} - \text{PERSONNE} \}$$

$$p_1 = p(\text{PERSONNE} - \text{utilisateur} - \text{TÉLÉCOM})$$

$$p_2 = p(\text{TÉLÉCOM} - \text{en contact} - \text{TÉLÉCOM})$$

$$p_3 = p(\text{PERSONNE} - \text{en contact} - \text{TÉLÉCOM})$$

$$p_4 = p(\text{TÉLÉCOM} - \text{utilisateur} - \text{PERSONNE})$$

$$P_{\text{lien direct}} = p_1 \cdot p_2 \cdot p_3 \cdot p_4$$

Si p_1 et p_2 et p_3 et $p_4 \geq 0.85$, alors $\text{alerte}(\text{lien direct}) = 1$,
sinon $\text{alerte}(\text{lien direct}) = 0$.

Figure 69 Illustration de la démarche suggérée pour contrôler les chaînes téléphoniques longues

Une solution similaire peut être envisagée en l'absence d'informations de dates et d'heures pour la gestion des chemins nécessitant de disposer de tels éléments sur tout ou partie

des symboles qui les composent. Pour les chemins contenant des liens de type « localisation » successifs (figure 70), il s’agirait alors de regarder si les attributs de date et d’heure existent et d’agir en fonction : si ces informations ne sont pas renseignées ou bien si elles ne correspondent pas, alors l’analyste préférera que le chemin n’induisse pas de lien direct ; si elles sont disponibles et correspondent, alors le lien direct sera généré et sa pondération calculée.

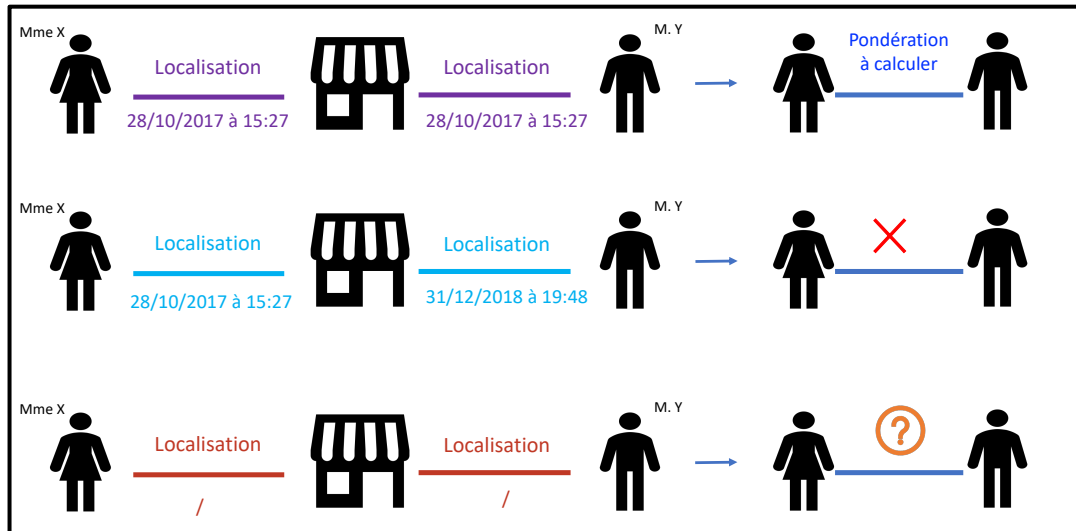


Figure 70 Prise en compte de la disponibilité et des informations de dates et heures

Cette idée pourrait être matérialisée comme indiqué sur la figure 71.

$L : \{ \text{PERSONNE} - \text{localisation} - \text{LIEU PUBLIC} - \text{localisation} - \text{PERSONNE} \}$

$p_1 = p(\text{PERSONNE} - \text{localisation} - \text{LIEU PUBLIC})$

$p_2 = p(\text{LIEU PUBLIC} - \text{localisation} - \text{PERSONNE})$

Si $\text{date}(\text{PERSONNE} - \text{localisation} - \text{LIEU PUBLIC})$, $\text{date}(\text{LIEU PUBLIC} - \text{localisation} - \text{PERSONNE})$, $\text{heure}(\text{PERSONNE} - \text{localisation} - \text{LIEU PUBLIC})$, et $\text{date}(\text{LIEU PUBLIC} - \text{localisation} - \text{PERSONNE})$ existent alors :

Si $\text{date}(\text{PERSONNE} - \text{localisation} - \text{LIEU PUBLIC})$

= $\text{date}(\text{LIEU PUBLIC} - \text{localisation} - \text{PERSONNE})$,

et $\text{heure}(\text{PERSONNE} - \text{localisation} - \text{LIEU PUBLIC})$

= $\text{heure}(\text{LIEU PUBLIC} - \text{localisation} - \text{PERSONNE})$,

alors $P_{\text{lien direct}} = p_1 \cdot p_2$

sinon $p_1 = p_2 = 0 \Rightarrow P_{\text{lien direct}} = 0$

sinon $p_1 = p_2 = 0 \Rightarrow P_{\text{lien direct}} = 0$

Figure 71 Illustration de la démarche suggérée pour gérer les dates et heures sur les liens de type « localisation »

Enfin, il faudrait donner à l'analyste la possibilité d'accéder et de contrôler facilement les *patterns* de chemins de *data0* ayant induit les liens directs d'*Investigation_network*. Cela permettrait, en ayant connaissance des limites énoncées, de mieux repérer les erreurs, et de les corriger manuellement en renseignant le fichier de liens à rejeter et en appliquant à *Investigation_network* le script de suppression des arêtes pour lesquelles l'analyste estime qu'elles n'existent pas sur le terrain.

En dernier point, il est important de préciser que ces solutions ont été imaginées par la doctorante. Leur mise en œuvre, non triviale, est au-delà de son champ de compétences, et leur faisabilité technique devra à l'avenir être discutée et évaluée avec Bruno Pinaud, auteur de l'algorithme.

c) Critères de collecte et de saisie de la donnée par les enquêteurs et analystes

Une méthode de saisie orientée vers l'utilisation de modèles automatisés permettrait d'obtenir des résultats d'une meilleure précision. L'objectif de la thèse étant de proposer une méthodologie de représentation, de visualisation et d'analyse de réseaux criminels sur la base des données telles que collectées et saisies lors des enquêtes de gendarmerie, il ne s'agit pas de proposer des critères de collecte et de saisie ne correspondant pas aux méthodes actuellement utilisées par la gendarmerie. Il s'agit d'incorporer à la matrice d'analyse des informations déjà existantes mais actuellement peu ou pas prises en compte, soit parce qu'elles n'apportent que peu d'éléments à l'enquête, soit parce qu'elles sont gérées différemment par les analystes et les enquêteurs. En effet, l'analyste criminel raisonne de manière continue lors de la saisie des données dans la matrice d'analyse, il va donc traiter cognitivement un grand nombre d'informations et les exploiter sans les saisir formellement. Pour pouvoir automatiser ces traitements et appuyer l'analyste dans son travail, toutes les informations devant être exploitées doivent être collectées et saisies de manière standardisée, homogène et uniforme. Ainsi, comme il a été discuté dans les sections précédentes, le modèle ne peut fonctionner de manière optimale que si les informations de dates et d'heures sont disponibles pour tous les liens ponctuels, tels que les localisations, les contacts, les échanges d'objets et d'argent ...

Nous proposons également, pour améliorer la précision et les performances du modèle, de mettre en place un nouveau critère de saisie prenant en compte l'avis de l'enquêteur sur l'implication potentielle de la structure dans les activités criminelles investiguées, *i.e.* si, à ce stade de l'enquête, il a ou non des raisons de penser que le commerce ou l'entreprise y prend

part (figure 72)⁴⁰⁰. L'ontologie des types d'entités ne serait alors plus basée uniquement sur la sémantique du libellé et l'icône⁴⁰¹, mais également sur ce nouvel attribut (tableau 60). En outre, la matrice de saisie ne permet actuellement de renseigner des attributs que pour les entités sources de lien et non pour les entités destinations. Il faudrait ainsi ajouter à la matrice des colonnes permettant de saisir également les informations relatives aux entités qui n'apparaissent que dans la colonne « ENTITÉ DESTINATION »⁴⁰².

ENTITE SOURCE	ICONE SOURCE	Taille	Implication	CATEGORIE_NOEUD_SOURCE	LIBELLE LIEN	CATEGORIE_LIEN	DESTINATION	ICONE DE	CATEGORIE_NOEUD_DESTINATION
LAVERIE DUPONT	Magasin				a comme gerant	professionnel	M. X	Homme	PERSONNE
			<ul style="list-style-type: none"> Lien établi Lien suspecté <li style="background-color: #f4a460;">Pas de lien 						

Figure 72 Prise en compte de l'avis expérimenté de l'enquêteur

Tableau 60 Ontologie basée sur la sémantique et sur l'implication potentielle des structures

ICONE SOURCE	Implication	TYPE ENTITE
Café	Lien établi	COMMERCE IMPLIQUÉ
	Lien suspecté	COMMERCE
	Pas de lien	COMMERCE NON IMPLIQUÉ
Restaurant	Lien établi	COMMERCE IMPLIQUÉ
	Lien suspecté	COMMERCE
	Pas de lien	COMMERCE NON IMPLIQUÉ
Magasin	Lien établi	COMMERCE IMPLIQUÉ
	Lien suspecté	COMMERCE
	Pas de lien	COMMERCE NON IMPLIQUÉ
Bureau	Lien établi	PERSONNE MORALE IMPLIQUÉE
	Lien suspecté	PERSONNE MORALE IMPLIQUÉE
	Pas de lien	PERSONNE MORALE NON IMPLIQUÉE

Les questions qui se posent lors de l'analyse sémantique de *patterns* de chemins contenant des lieux publics mettent en évidence que le modèle n'est pas satisfaisant pour traiter ces chemins de manière générique. En effet, le nombre de variables à prendre en compte pour décider sémantiquement si le *pattern* doit être accepté ou rejeté augmente considérablement avec la taille et le type de lieu public en question. Prenons à titre d'exemple le cas apparemment simple de la figure 73.

⁴⁰⁰ Ce critère pourra évoluer au-fur-et-à-mesure des avancées de l'investigation (dans une logique de soutien à l'enquête) et des analyses (dans une logique d'analyse de phénomènes).

⁴⁰¹ Cf. section IV.1 c) « Standardisation du contenu par la création d'ontologies sémantiques : préparation aux traitements automatisés ».

⁴⁰² Cf. section III.2 c) « Matrices d'analyse : la donnée structurée », tableaux 3 à 5.

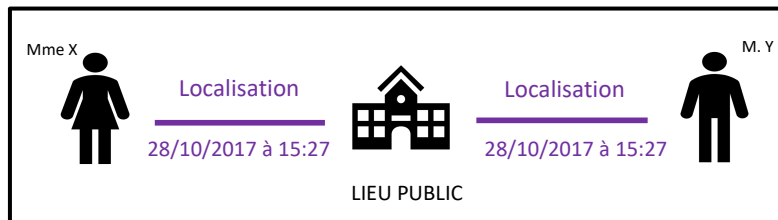


Figure 73 Exemple de situation problématique

Dans cette situation, Mme X et M. Y sont localisés au même endroit, à la même heure. La première question qui émerge lors de l'analyse d'un tel pattern est « de quel type de lieu public s'agit-il ? ». La gestion du *pattern* sera très différente s'il s'agit d'un petit salon de coiffure où s'il s'agit d'un aéroport international. Les informations sous-jacentes dont l'analyste souhaite disposer lorsqu'il s'interroge sur le type de lieu sont notamment la capacité d'accueil, les flux de personnes, la nature des échanges qui s'y déroulent.

Le modèle proposé a une vocation d'aide à la décision pour l'enquêteur et/ou l'analyste et doit être adaptable (paramétrable) selon les préoccupations de chacun. Dans certaines situations impliquant des lieux publics, il doit cependant être complété par une discrimination humaine ; d'une part car la décision finale reviendra toujours à l'analyste, d'autre part car les facteurs de discrimination appliqués par l'humain viennent enrichir le modèle lui-même.

Nous ne remettons pas ici en cause les compétences de l'enquêteur pour la collecte des éléments ni de l'analyste pour la saisie des données : leur expérience, leurs connaissances et leurs raisonnements sont indispensables à la qualité de l'enquête et de l'analyse. Nous proposons de stocker et de traiter les informations de manière différente des méthodes actuelles tout en conservant les procédés de collecte et de saisie de la donnée. Nous mettons en évidence le fait que la prise en compte des informations disponibles et exploitables lors de l'extraction automatisée nécessite leur saisie formelle dans le fichier d'entrée du modèle développé (la matrice d'analyse). Ces informations pourraient alors dans un premier temps être prises en compte lors de la catégorisation des entités et l'attribution des forces liantes des symboles minimaux, puis, dans un second temps, par l'algorithme lors des fusions et agrégations de liens, si des développements futurs étaient envisagés.

Les performances et taux d'erreurs de l'algorithme étant à présent estimés et les limites du modèle explicitées et discutées, nous pouvons procéder aux analyses de réseaux sociaux sur le sociogramme *Investigation_network* obtenu par filtrage de *Social_network*.

CHAPITRE VII ANALYSE DE RÉSEAUX SOCIAUX : APPLICATION AU SOCIOGRAMME *INR*

Comme discuté au chapitre précédent, le seuil de filtrage dépend des objectifs et besoins de l'analyste ainsi que de la logique de renseignement criminel dans laquelle elle se place. Dans cette thèse, nous adoptons une démarche d'analyse de phénomènes criminels et de profils de criminalité organisée. L'objectif est de proposer une méthodologie de représentation, de modélisation et d'analyse de réseaux criminels appuyant les analystes en renseignement criminel et les analystes criminels dans la détection de fonctions et positions stratégiques pour les organisations criminelles. Ainsi, nous travaillons à présent sur le sociogramme *Investigation_network_{rens}* (*INR*) obtenu par filtrage du graphe pondéré *Social_network* avec la valeur seuil $S_{Rens} = 0.13$ (figure 74).

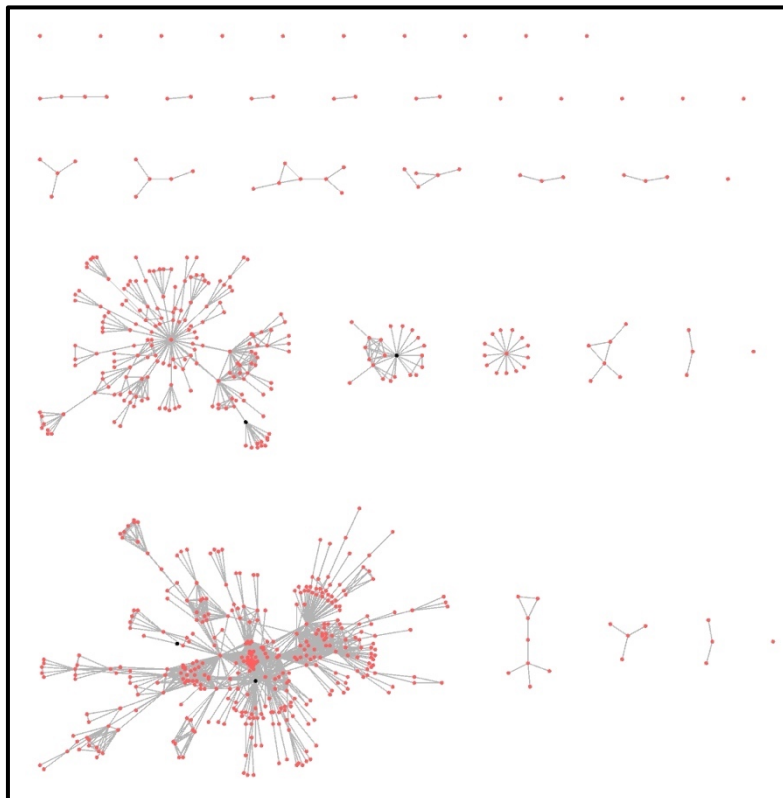


Figure 74 Sociogramme *INR*

L'idée d'utiliser les concepts et mesures issues de l'ARS pour l'analyse criminelle de réseaux a été émise il y a plus de 30 ans. A l'inverse des approches reposant sur des hypothèses et/ou affirmations sur les fonctionnements dans un contexte criminel donné, l'ARS propose une méthode de recherche et d'évaluation de structure. L'approche réseau permet en effet d'une part de gérer les données complexes (par leur nombre élevé d'acteurs et de relations) que sont les

données issues d'actes d'enquête⁴⁰³, et d'autre part de les visualiser et de les analyser au regard du système global que définit un réseau⁴⁰⁴. En particulier, de nombreuses études de cas ont montré, en utilisant l'ARS, que les réseaux criminels sont très flexibles, d'où leur résilience face aux interventions des forces de l'ordre⁴⁰⁵.

VII.1 Propriétés du sociogramme *INR*

a) Traitements préliminaires à l'analyse

Cette nouvelle phase d'analyse s'inscrit également dans le cycle du renseignement. Ce dernier étant itératif, les résultats de chaque phase font l'objet d'un pré-traitement avant de passer à la phase d'analyse suivante, ici l'analyse du réseau social. Le sociogramme *INR* est donc dans un premier temps nettoyé pour ne conserver que les éléments d'intérêt exploitables.

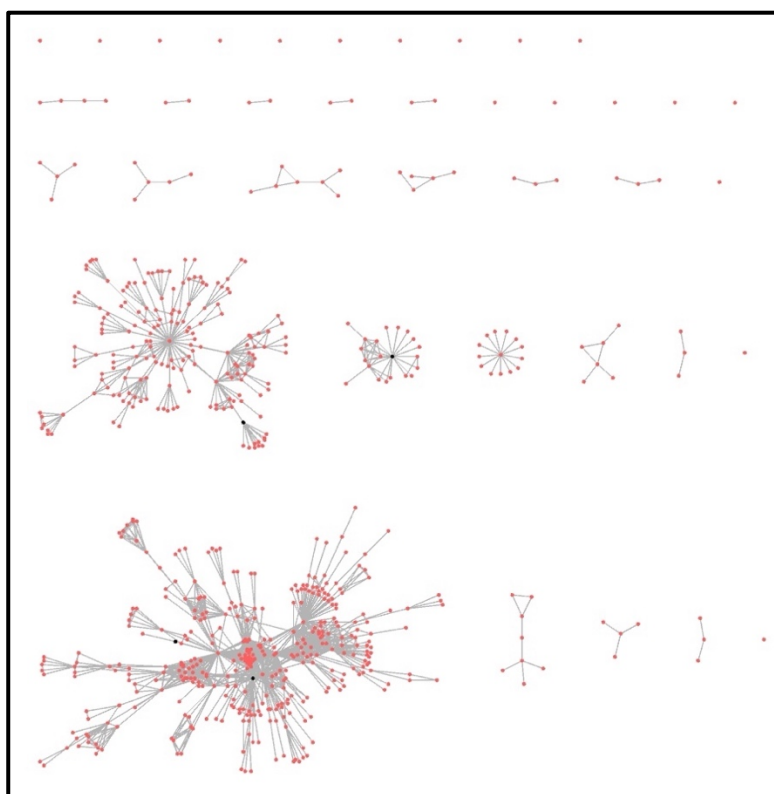


Figure 75 Sociogramme *INR*

⁴⁰³ Malcolm K. Sparrow, « The application of network analysis to criminal intelligence: An assessment of the prospects », *Social Networks*, vol. 13, n° 3, 1991, pp. 251-274 ; Peter Klerks, « The Network Paradigm Applied to Criminal Organisations: Theoretical nitpicking or a relevant doctrine for investigators? Recent developments in the Netherlands », *Connections*, vol. 24, n° 1, 2001, pp. 53-65.

⁴⁰⁴ Carlo Morselli, « Assessing Vulnerable and Strategic Positions in a Criminal Network », *op. cit.*

⁴⁰⁵ Geoffrey Pearson et Dick Hobbs, *Middle Market Drug Distribution: Home Office Research Study*, Home Office, 2001, 91 p ; Ahmet Icduygu et Sule Toktas, « How do Smuggling and Trafficking Operate via Irregular Border Crossings in the Middle East? Evidence from Fieldwork in Turkey », *International Migration*, vol. 40, n° 6, 12/2002, pp. 25-54 ; J. M. McGloin, « Policy Intervention Considerations of a Network Analysis of Street Gangs », *Criminology & Public Policy*, vol. 4, n° 1, 2005, pp. 607-635 ; Carlo Morselli, *Inside criminal networks*, New York, NY, Springer, 2009, 203 p.

Le sociogramme INR est dit « déconnecté », c'est-à-dire qu'il n'existe pas de chemins entre toutes les paires de nœuds⁴⁰⁶. Les nœuds sont séparés en plusieurs groupes distincts, appelés « composantes connexes ». *INR* contient initialement 38 composantes connexes.

L'objectif étant d'étudier les interactions sociales à l'intérieur de l'organisation criminelle, les 18 nœuds isolés⁴⁰⁷ ont été supprimés, dans la mesure où ils ne sont dans cette optique pas informatifs⁴⁰⁸. Pour les mêmes raisons, les 11 composantes du graphe contenant $N < 5$ personnes ont été supprimées. Comme précisé dans la description de l'affaire⁴⁰⁹, les services proposés par le réseau couvraient tant le transfert d'argent de clients particuliers à l'étranger *via* un système financier parallèle (dit de « l'*hawala* ») que le blanchiment massif d'argent pour des clients tels que des réseaux de trafic de stupéfiants. Nous estimons, sur la base des informations dont nous disposons, que ces composantes de petite taille font vraisemblablement référence à des clients particuliers du réseau (et à leur entourage). Ces derniers auraient bénéficié ponctuellement des services de blanchiment et/ou de transfert de fonds à l'étranger, mais ne seraient pas impliqués dans le réseau criminel⁴¹⁰. Ainsi, *INR* est réduit de 38 à 9 composantes connexes.

⁴⁰⁶ Stanley Wasserman et Katherine Faust, *Social Network Analysis Methods and Applications*, *op. cit.*

⁴⁰⁷ Les nœuds de degré 0, *i.e.* les personnes qui ne sont connectées à aucun autre individu du graphe.

⁴⁰⁸ Nous n'affirmons pas que ces données ne sont pas pertinentes, mais plutôt qu'elles ne sont pas exploitables en ARS pour étudier les relations et les échanges sociaux qui ont lieu dans l'organisation criminelle. D'autres analyses pourraient cependant révéler des éléments intéressants, tels que des rôles partiels tenus par des acteurs de manière ponctuelle, ou maintenus volontairement à distance de l'organisation criminelle dans son ensemble.

⁴⁰⁹ Cf. chapitre III « Le cas d'usage et les données d'étude ».

⁴¹⁰ Nous n'affirmons pas que ces données ne sont pas pertinentes, mais plutôt qu'elles ne sont pas exploitables en ARS pour étudier les relations et les échanges sociaux qui ont lieu dans l'organisation criminelle. D'autres analyses pourraient cependant révéler des éléments intéressants, notamment sur le volume de clients particuliers du réseau par rapport aux « clients *business* » (les organisations criminelles qui utilisent les services de blanchiment de manière récurrente et à une échelle financière bien plus large).

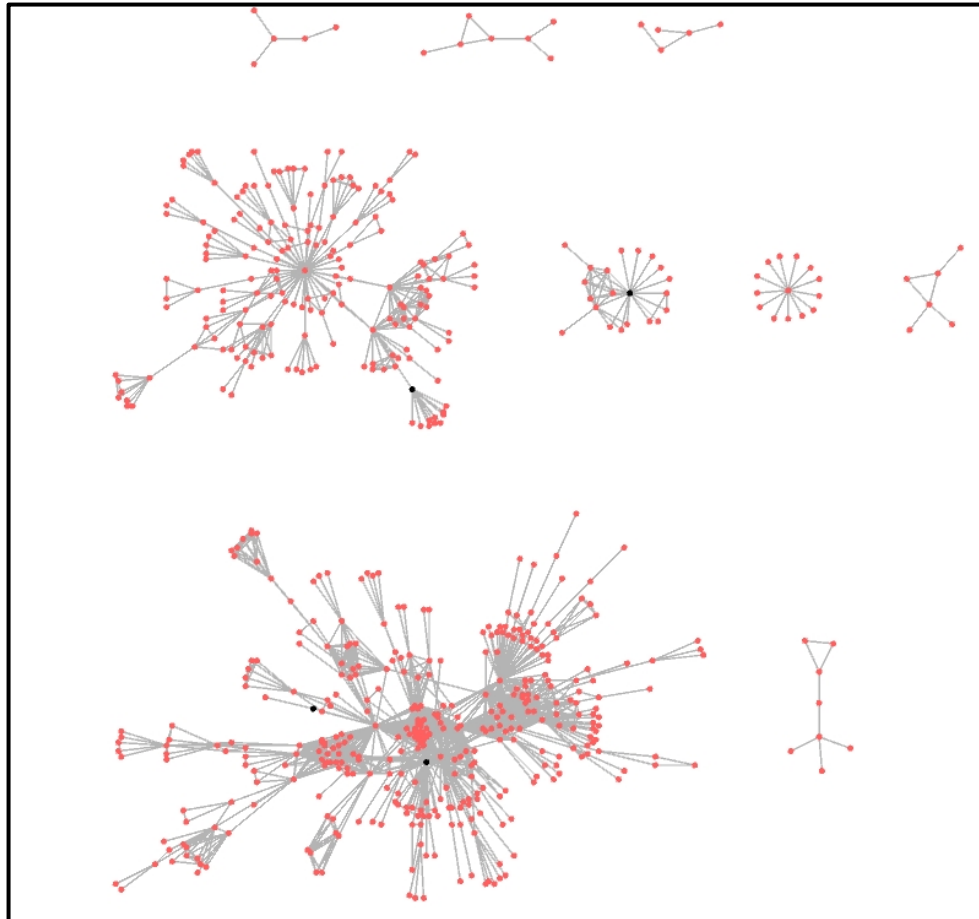


Figure 76 Sociogramme *INR* pré-traité

b) Description du graphe final : propriétés structurelles

INR contient $N = 615$ nœuds pour $L = 2\,065$ arêtes. Son diamètre est de 9, ce qui signifie que les deux individus les plus éloignés doivent passer par au moins 8 acteurs intermédiaires pour communiquer. Sa densité, calculée en rapportant le nombre L d'arêtes du graphe sur le nombre d'arêtes possible⁴¹¹ est égale à 0.011. Cette valeur est extrêmement faible, indiquant qu'il y a peu de liens par rapport au nombre de nœuds : les individus sont de manière générale assez peu connectés les uns aux autres.

$$d = \frac{2L}{N(N-1)}$$

Chacune des 9 composantes du graphe comporte au moins un *cutpoint* et un *bridge*. Un nœud v_i est un *cutpoint* si sa suppression du graphe entraîne l'augmentation du nombre de

⁴¹¹*i.e.* le nombre d'arêtes que contiendrait le graphe si toutes les paires de nœuds étaient connectées, $\frac{N(N-1)}{2}$.

composantes. De même, une arête est un *bridge* si sa suppression entraîne l'augmentation du nombre de composantes (figure 77).

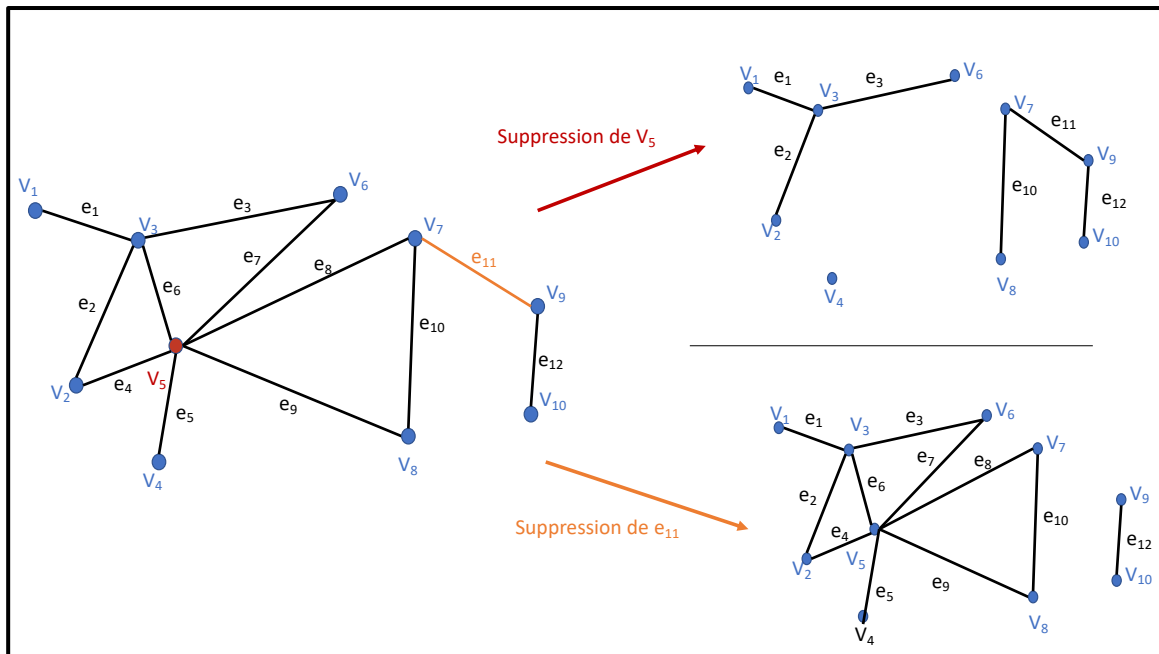


Figure 77 Le nœud V_5 est un cutpoint et l'arête e_{11} est un bridge

De manière générale, et en particulier dans les graphes représentant plusieurs centaines d'entités et de relations (tels qu'*INR*), les concepts de *cutpoint* et de *bridge* peuvent être étendus à un ensemble de nœuds (et d'arêtes) nécessaires au maintien de la connexion du graphe. Typiquement, sur la figure 78, le retrait des deux nœuds orange et vert (et par conséquent des arêtes associées) entraîne une fragmentation de la composante S en la découpant en 9 nouvelles composantes, dont 3 sont dites « triviales », c'est-à-dire qu'elles ne comportent qu'un seul nœud et aucune arête.

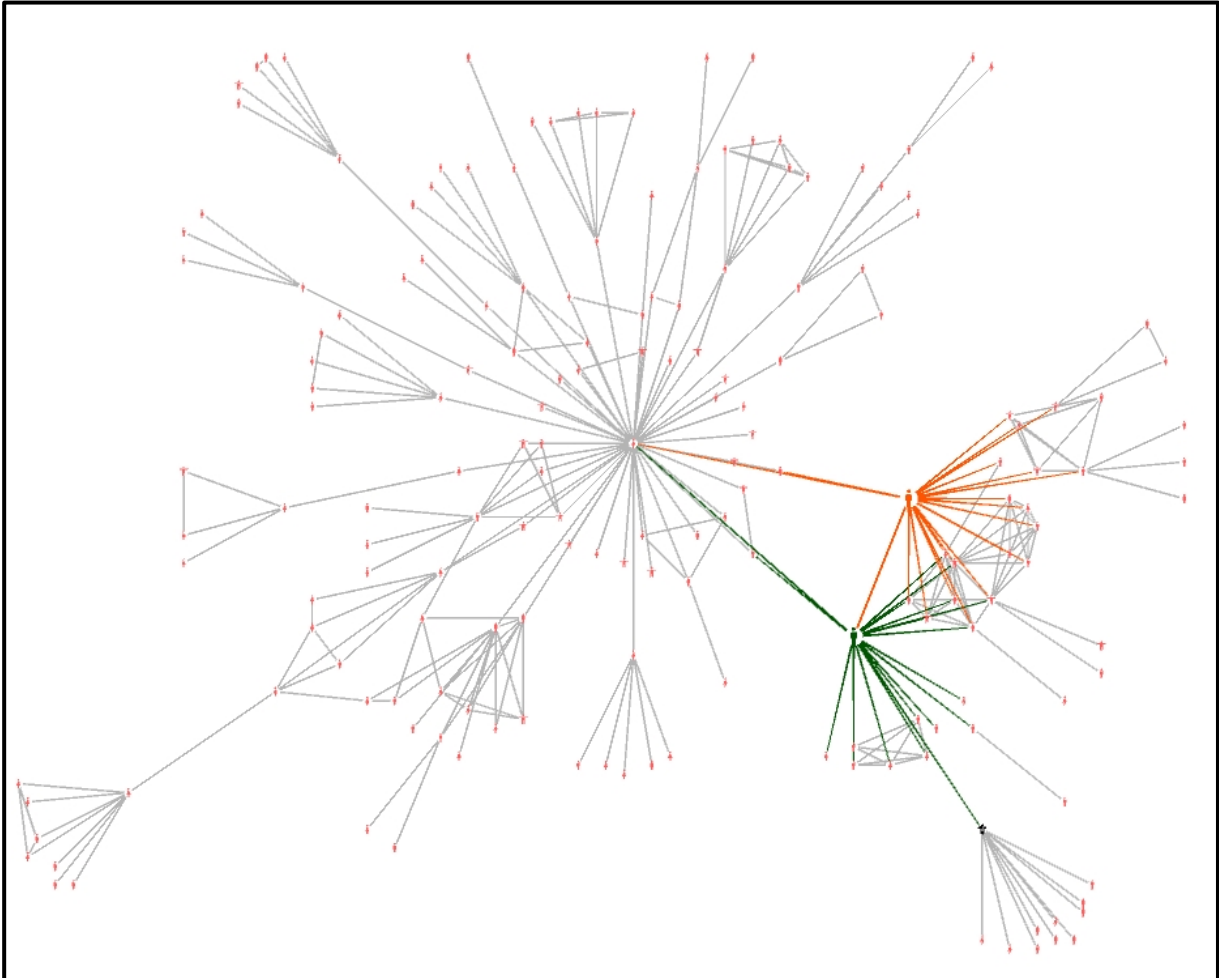


Figure 78 Exemple d'un ensemble de deux nœuds dont la suppression simultanée entraîne une fragmentation de S .

La connectivité d'un graphe est l'une des manières de mesurer son niveau de cohésion ou sa robustesse, dans la mesure où un graphe non cohésif a une structure vulnérable à la suppression de nœuds et/ou d'arêtes. Les caractéristiques générales d'un graphe cohésif sont une forte densité, beaucoup de nœuds avec une centralité de degré importante, de nombreux chemins entre paires de nœuds, des chemins relativement courts et un faible diamètre⁴¹². Deux indicateurs en particulier permettent de définir la connectivité d'un graphe. Le premier, la *node-connectivity* (κ), est basé sur les effets de la suppression de nœuds ; le second, l'*edge-connectivity* (λ), est basé sur les effets de la suppression d'arêtes. Dans un graphe G (ou une composante connexe G_s d'un graphe déconnecté), κ et λ sont le nombre minimal de nœuds et d'arêtes devant être retirés pour déconnecter G (ou G_s)⁴¹³. Dans le cas du sociogramme *INR*, ses 9 composantes connexes G_{s_i} possédant au moins un *cutpoint* et un *bridge*, $\kappa_{s_i} = 1$ et $\lambda_{s_i} = 1 \forall i \in [1 ; 9]$ ⁴¹⁴.

⁴¹² Stanley Wasserman et Katherine Faust, *Social Network Analysis Methods and Applications*, op. cit.

⁴¹³ Frank Harary et Robert Z. Norman, *Graph Theory as a Mathematical Model in Social Science*, op. cit.

⁴¹⁴ i.e. pour tout i allant de 1 à 9, les indices des composantes connexes de *INR*.

INR est donc très peu dense et déconnecté ; ses composantes connexes sont de faible connectivité, indiquant que le graphe est vulnérable à la suppression de nœuds et d'arêtes (tableau 61). Il s'agit maintenant d'identifier les nœuds – ou groupes de nœuds– sur lesquels repose sa structure dans le but de détecter les positions et fonctions importantes pour l'organisation criminelle.

Tableau 61 Caractéristiques du graphe et de ses 9 composantes connexes

Composante	Nombre de nœuds	Nombre d'arrêtes	Diamètre	Densité	Node-connectivity	Edge-connectivity
G_{S_1}	$N_{S_1} = 5$	$L_{S_1} = 4$	3	0.4	$\kappa_{S_1} = 1$	$\lambda_{S_1} = 1$
G_{S_2}	$N_{S_2} = 7$	$L_{S_2} = 7$	4	0.33	$\kappa_{S_2} = 1$	$\lambda_{S_2} = 1$
G_{S_3}	$N_{S_3} = 5$	$L_{S_3} = 4$	3	0.4	$\kappa_{S_3} = 1$	$\lambda_{S_3} = 1$
G_{S_4}	$N_{S_4} = 181$	$L_{S_4} = 306$	7	0.019	$\kappa_{S_4} = 1$	$\lambda_{S_4} = 1$
G_{S_5}	$N_{S_5} = 21$	$L_{S_5} = 40$	3	0.19	$\kappa_{S_5} = 1$	$\lambda_{S_5} = 1$
G_{S_6}	$N_{S_6} = 14$	$L_{S_6} = 13$	2	0.17	$\kappa_{S_6} = 1$	$\lambda_{S_6} = 1$
G_{S_7}	$N_{S_7} = 6$	$L_{S_7} = 6$	3	0.4	$\kappa_{S_7} = 1$	$\lambda_{S_7} = 1$
G_{S_8}	$N_{S_8} = 368$	$L_{S_8} = 1\ 677$	9	0.025	$\kappa_{S_8} = 1$	$\lambda_{S_8} = 1$
G_{S_9}	$N_{S_9} = 8$	$L_{S_9} = 8$	4	0.29	$\kappa_{S_9} = 1$	$\lambda_{S_9} = 1$
G	$N = 615$	$L = 2\ 065$	9	0.011	$\kappa = 1$	$\lambda = 1$

VII.2 Analyse du réseau social : propriétés sociométriques des nœuds

L'un des premiers usages de la théorie des graphes en ARS est l'identification des acteurs « les plus importants » d'un réseau social⁴¹⁵. L'« importance » d'un nœud est envisagée en termes de contribution à la structure du réseau : c'est la notion de centralité, englobant une famille de concepts. En effet, la notion « d'importance » n'est pas universelle, et peut se définir de plusieurs manières, selon les critères que considère l'analyste. Un nœud peut être structurellement central selon plusieurs considérations⁴¹⁶ : soit car il permet (ou accélère) la

⁴¹⁵ Stanley Wasserman et Katherine Faust, *Social Network Analysis Methods and Applications*, op. cit.

⁴¹⁶ Stephen P. Borgatti, Martin G. Everett et Jeffrey C. Johnson, *Analyzing social networks*, op. cit.

connexion entre plusieurs parties du réseau⁴¹⁷, soit parce qu'un grand nombre d'arêtes lui sont incidentes, ou encore parce qu'il a une distance relativement courte avec l'ensemble des autres nœuds du graphe. Il existe donc une variété de mesures créées pour évaluer différentes formes d'importance et distinguer les protagonistes les plus importants.

a) Centralité de degré

Le degré de centralité⁴¹⁸ d'un nœud est calculé par simple comptage du nombre d'autres nœuds auxquels il est connecté⁴¹⁹. Bien que ce nombre en soi donne peu d'information, il permet d'étudier l'importance relative des nœuds les uns par rapport aux autres dans un graphe. En effet, un acteur avec un fort degré de centralité a un potentiel d'interaction directe avec les autres plus important⁴²⁰. On parle également d'« activité » dans le réseau.

Soit $G = (V, E)$ un graphe non orienté d'ordre N représenté par sa *sociomatrix* A ⁴²¹. La centralité de degré $C_{degré}(v_i)$ du nœud $v_i \in V$ est définie comme suit⁴²² :

$$C_{degré}(v_i) = \sum_j a_{ij}$$

La valeur maximale de la centralité de degré est obtenue pour un nœud directement connecté à tous les autres sommets du graphe, et vaut alors $N - 1$. Cette valeur est ainsi dépendante de la taille du graphe. Afin de standardiser cette mesure pour la rendre comparable d'un graphe à l'autre, Stanley Wasserman et Katherine Faust⁴²³ proposent de mesurer la proportion⁴²⁴ des sommets du graphe auxquels le nœud v_i est connecté plutôt que d'en compter le nombre, en divisant cette valeur par son maximal possible, $N - 1$:

$$C'_{degré}(v_i) = \frac{1}{N - 1} \sum_j a_{ij}$$

⁴¹⁷ Tel un *cutpoint*, seul connecteur entre deux composantes d'un graphe, mais également en permettant de diminuer la distance existante entre des nœuds éloignés en son absence.

⁴¹⁸ Dans un souci de lisibilité et par abus de langage, nous emploierons également le terme « degré » pour désigner la centralité de degré.

⁴¹⁹ Alain Degenne et Michel Forsé, *Les réseaux sociaux*, op. cit.

⁴²⁰ Paola Tubaro, *International encyclopedia of the social & behavioral sciences*, op. cit. ; Linton C. Freeman, « Centrality in social networks conceptual clarification », op. cit.

⁴²¹ Voir figure 13.

⁴²² Stanley Wasserman et Katherine Faust, *Social Network Analysis Methods and Applications*, op. cit.

⁴²³ *Ibid.*

⁴²⁴ Les valeurs standardisées étant égales aux valeurs non standardisées pondérées d'un facteur $\frac{1}{N-1}$, ces deux grandeurs sont proportionnelles.

La figure 79 représente le graphe *INR* en pondérant la taille des nœuds par leur centralité de degré.

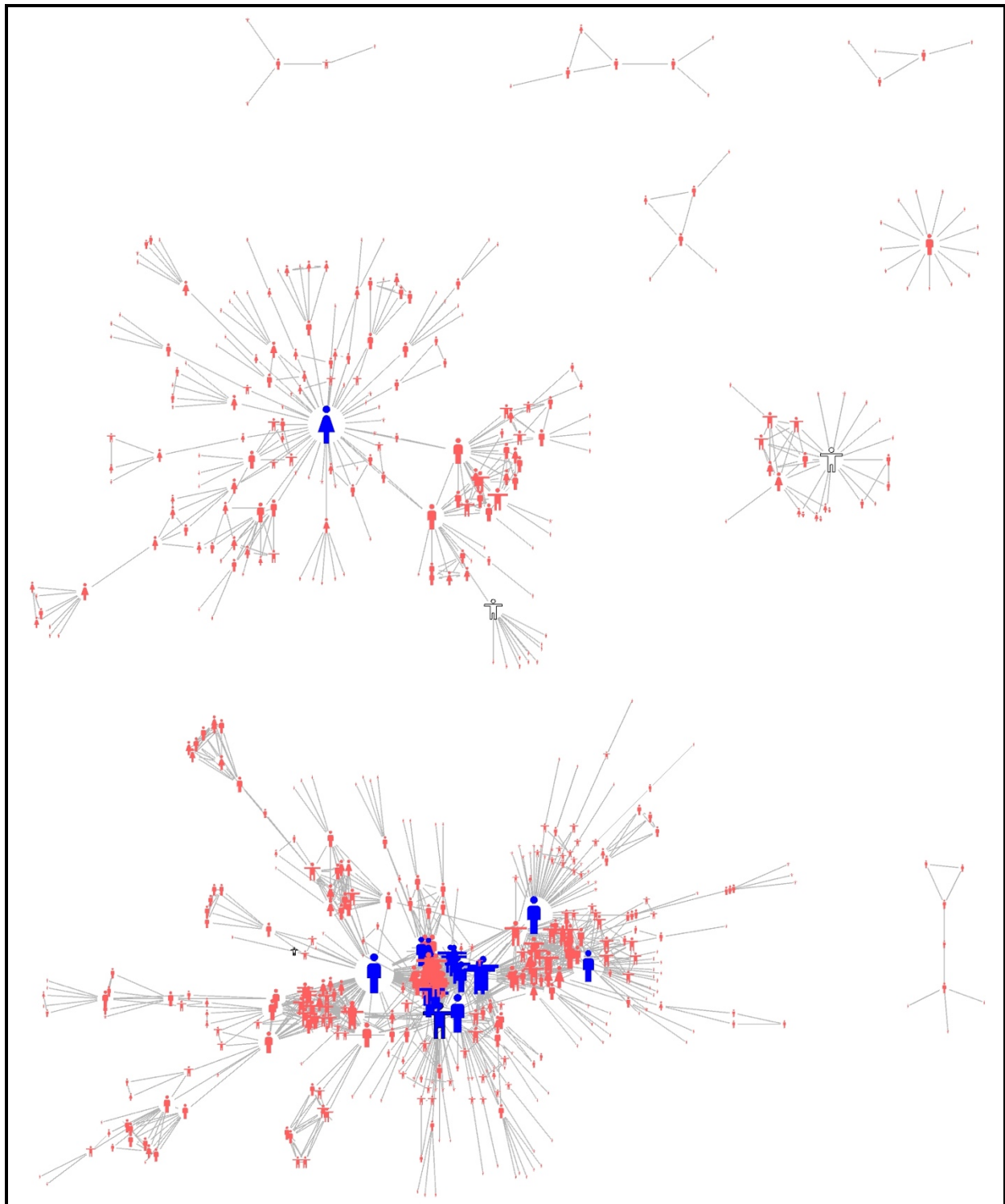


Figure 79 Pondération de la taille des nœuds par leur centralité de degré. Les nœuds bleus sont les 20 acteurs ayant les plus forts degrés.

Le tableau 62 reporte, pour les 20 premiers acteurs mis en évidence, les valeurs des centralités de degré ainsi que les valeurs standardisées (en pourcentage du réseau auquel accède directement un acteur).

Tableau 62 Nombre $C_{\text{degré}}(v_i)$ et proportion $C'_{\text{degré}}(v_i)$ de voisins directs des 20 nœuds avant les degrés les plus élevés, classés par centralité décroissante

ID_anonym	Degree	Degree_norm
1	83	13,52%
250	78	12,70%
348	74	12,05%
110	68	11,07%
345	62	10,10%
344	62	10,10%
402	50	8,14%
1115	49	7,98%
288	48	7,82%
113	45	7,33%
272	43	7,00%
265	39	6,35%
259	39	6,35%
319	39	6,35%
358	39	6,35%
292	38	6,19%
318	38	6,19%
356	38	6,19%
1684	37	6,03%
298	37	6,03%

Le nœud le plus connecté (ayant le plus de connexions directes) accède ainsi directement à 13,52% des nœuds du graphe. Seuls 6 acteurs, sur les 615 du réseau peuvent atteindre directement 10% ou plus de l'ensemble des individus impliqués. La centralité de degré moyenne est de 7 : en moyenne, un nœud du graphe est connecté avec 7 individus (soit 1,14% de l'effectif).

La figure 80 illustre la distribution et la répartition des valeurs de la centralité de degré pour l'ensemble des acteurs ; le tableau 63 en reporte les indicateurs statistiques de position et de dispersion. Pour mieux saisir leur sens, ces derniers sont, comme les visualisations, donnés pour les valeurs de centralité de degré non standardisées (*i.e.* en nombre de connexions directes).

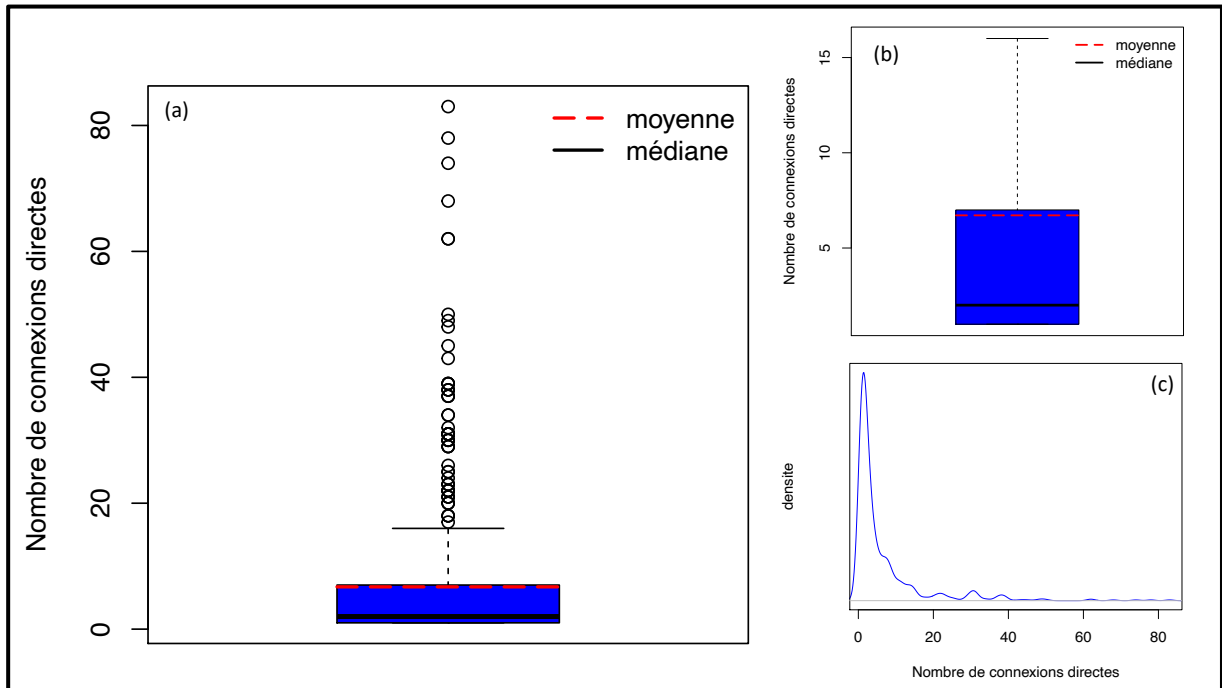


Figure 80 Répartition des valeurs de centralité de degré (standardisée). (a) *Boxplot* ; (b) *Boxplot* sans les *outliers* ; (c) Courbe de densité

Le graphique (a) est un *boxplot* représentant la répartition des valeurs de la centralité de degré standardisée. Le bas du rectangle (ou boîte) bleu représente le 1er quartile, *i.e.* la valeur de centralité en dessous de laquelle se trouvent 25% des acteurs ; le haut du rectangle bleu représente le 3^{ème} quartile, valeur de centralité en deçà de laquelle se trouvent 75% des acteurs. La médiane, qui représente la valeur en deçà de laquelle se trouvent 50% des acteurs, est matérialisée par le segment noir, et la moyenne par les pointillés rouges. Les *outliers*, ou valeurs extrêmes (par rapport à la distribution de l'ensemble des valeurs) sont représentées par les cercles noirs. Le graphique (b) est également un *boxplot* des valeurs de la centralité de degré standardisée, duquel ont été retirées les valeurs extrêmes afin de mieux visualiser la répartition des valeurs là où elles sont concentrées. Le graphique (c) représente la densité de la centralité de degré sur l'ensemble des 615 acteurs.

Tableau 63 Indicateurs de position et de dispersion pour la centralité de degré

Indicateur	Valeur	Signification
Moyenne	7	En moyenne, un nœud est directement connecté à 7 acteurs (soit à 1,14% de l'effectif).
Écart-type	10,8	La dispersion moyenne autour de la moyenne est de 10,8, c'est-à-dire qu'en moyenne, les acteurs ont une centralité de degré qui s'écarte de plus ou moins 10,8 connexions directes du nombre moyen de connexions ⁴²⁵ . Cet écart-type est très élevé en comparaison de la moyenne. La valeur minimale étant de 1, les valeurs ne s'écartent pas de la moyenne vers le bas. Cela signifie qu'il existe des valeurs de centralité de degré très élevées par rapport au nombre moyen de 7 contacts directs.
1^{er} quartile Q1	1	25% des acteurs sont directement connectés à 1 individu.
Médiane	2	50% des acteurs sont directement connectés à 2 individus ou moins.
3^{ème} quartile Q3	7	75% des acteurs sont directement connectés à 7 individus ou moins.
Écart interquartile (IQR = Q3 – Q1)	6	Les valeurs extrêmes sont celles supérieures à $Q3 + 1,5 \cdot IQR = 16$ connexions directes. Les acteurs ayant plus de 16 connexions directes (connectés à plus de 2,6% des nœuds) sont, par rapport à l'ensemble des acteurs, considérés comme extrêmement connectés. Cela concerne 63 individus, soit 10,2% de l'effectif ⁴²⁶ .
Min	1	Les acteurs sont tous connectés à au moins un individu.
Max	83	L'acteur le plus connecté est en relation avec 83 individus.

Ainsi, la plupart des acteurs sont caractérisés par une centralité de degré très faible, tandis qu'un petit nombre d'entre eux ont des valeurs extrêmement élevées par rapport à l'ensemble de la distribution. Bien que les acteurs soient assez faiblement connectés, la mesure de la centralité de degré permet de mettre en évidence que 10% d'entre eux ont un fort degré de connexion.

b) Centralité de proximité

La centralité de proximité est une mesure de centralité globale, c'est-à-dire prenant en compte l'ensemble des sommets d'un graphe connecté (ou d'une composante d'un graphe déconnecté), à la différence de la centralité de degré qui ne tient compte que des sommets

⁴²⁵ Une centralité ne pouvant être négative ou nulle (dans la mesure où tous les acteurs isolés ont été supprimés), la valeur basse est de 1.

⁴²⁶ Le tableau des centralités de degré de ces 62 acteurs est disponible en annexe 15.

adjacents au nœud v_i . Elle permet de déterminer si un nœud v_i est globalement proche de l'ensemble des autres nœuds qu'il peut atteindre (*i.e.* du graphe G connecté ou de la composante connexe S à laquelle v_i appartient) et s'obtient en calculant sa proximité moyenne vis-à-vis des autres nœuds⁴²⁷. En pratique, il s'agit de l'inverse de la somme des longueurs des chemins géodésiques⁴²⁸ entre le nœud d'intérêt v_i et chacun des autres nœuds⁴²⁹.

Plus sa valeur est importante pour un nœud donné, plus il a la possibilité d'accéder à un grand nombre d'autres nœuds *via* des trajets courts. Ainsi, une centralité de proximité élevée pour un acteur est un signe d'indépendance⁴³⁰ : il peut facilement recevoir ou passer de l'information, et accède rapidement aux communications et ressources qui passent par le réseau, sans avoir besoin de s'appuyer sur beaucoup d'autres membres.

Soient $G = (V, E)$ un graphe non orienté d'ordre N et $d(v_i, v_j)$ la distance entre les sommets v_i et v_j . La centralité de proximité $C_{proximité}(v_i)$ du nœud $v_i \in V$ est définie comme suit⁴³¹ :

$$C_{proximité}(v_i) = \frac{1}{\sum_j d(v_i, v_j)}$$

A l'inverse des centralités de degré ou d'intermédiarité, cet indicateur n'a de sens qu'à l'intérieur d'un graphe connecté, ou d'une composante d'un graphe déconnecté⁴³². En effet, par définition, il n'existe pas de chemins reliant des acteurs appartenant à des composantes connexes distinctes S_1 et S_2 : la distance entre $v_i \in S_1$ et $v_j \in S_2$ est donc infinie (figure 81).

⁴²⁷ Nacim Fateh Chikhi, *Calcul de centralité et identification de structures de communautés dans les graphes de documents*, *op. cit.*

⁴²⁸ La plus courte distance entre deux nœuds est la longueur du chemin géodésique, ou plus court chemin, les reliant.

⁴²⁹ Alain Degenne et Michel Forsé, *Les réseaux sociaux*, *op. cit.*

⁴³⁰ Michel Lacroix, « Littérature, analyse de réseaux et centralité », *op. cit.*

⁴³¹ Linton C. Freeman, « Centrality in social networks conceptual clarification », *op. cit.* ; Stanley Wasserman et Katherine Faust, *Social Network Analysis Methods and Applications*, *op. cit.*

⁴³² Stanley Wasserman et Katherine Faust, *Social Network Analysis Methods and Applications*, *op. cit.*

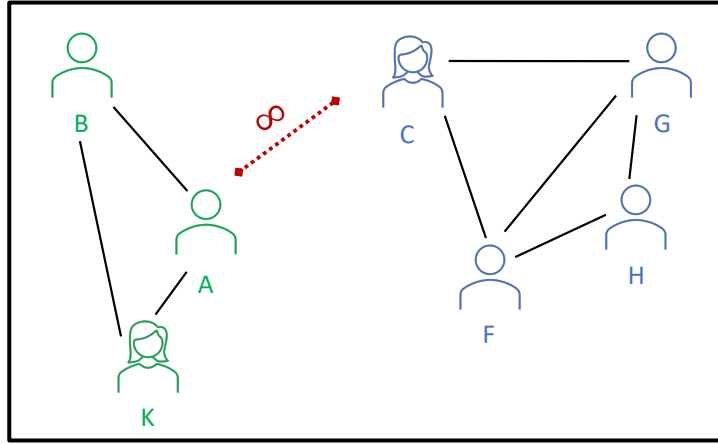


Figure 81 Les nœuds de la composante verte sont à une distance infinie des nœuds de la composante bleue (et réciproquement)

Ainsi, dans le cas d'un graphe déconnecté, la valeur de cet indicateur pour $v_i \in S_1$ ne peut être standardisée sur l'effectif total N du graphe, mais doit l'être sur le nombre de nœuds N_{S_1} de la composante S_1 . La centralité de proximité standardisée représente l'inverse de la distance moyenne entre le nœud v_i et tous les autres nœuds de sa composante, et se calcule comme suit :

$$C'_{proximité}(v_i) = \frac{N - 1}{\sum_j d(v_i, v_j)}$$

Pour un graphe connecté

$$C'_{proximité}(v_i) = \frac{N_s - 1}{\sum_j d(v_i, v_j)}$$

Pour une composante S d'un graphe déconnecté

Le graphe *INR* comporte 9 composantes connexes contenant de 5 à 368 nœuds. Les valeurs standardisées de la centralité de proximité ne sont donc pas proportionnelles aux valeurs non standardisées, dans la mesure où le facteur de pondération diffère selon les composantes. La standardisation permet ici de comparer les proportions des acteurs de leur composante que peuvent atteindre rapidement les différents nœuds.

Les figures 82 et 83 représentent respectivement le graphe *INR* en pondérant la taille des nœuds par leur centralité de proximité et par leur centralité de proximité standardisée. Les 20 nœuds dont la centralité est la plus élevée diffèrent légèrement selon la standardisation ou non de la mesure.

Dans le cas du graphe *INR*, déconnecté et dont les effectifs des différentes composantes connexes sont très hétérogènes, cette mesure s'avère peu informative. En effet, cet indicateur est fortement dépendant de la taille des composantes connexes ; avec un tel différentiel de taille,

la centralité de proximité favorisera toujours les nœuds des petites composantes par rapport à ceux des grandes composantes (figure 82).

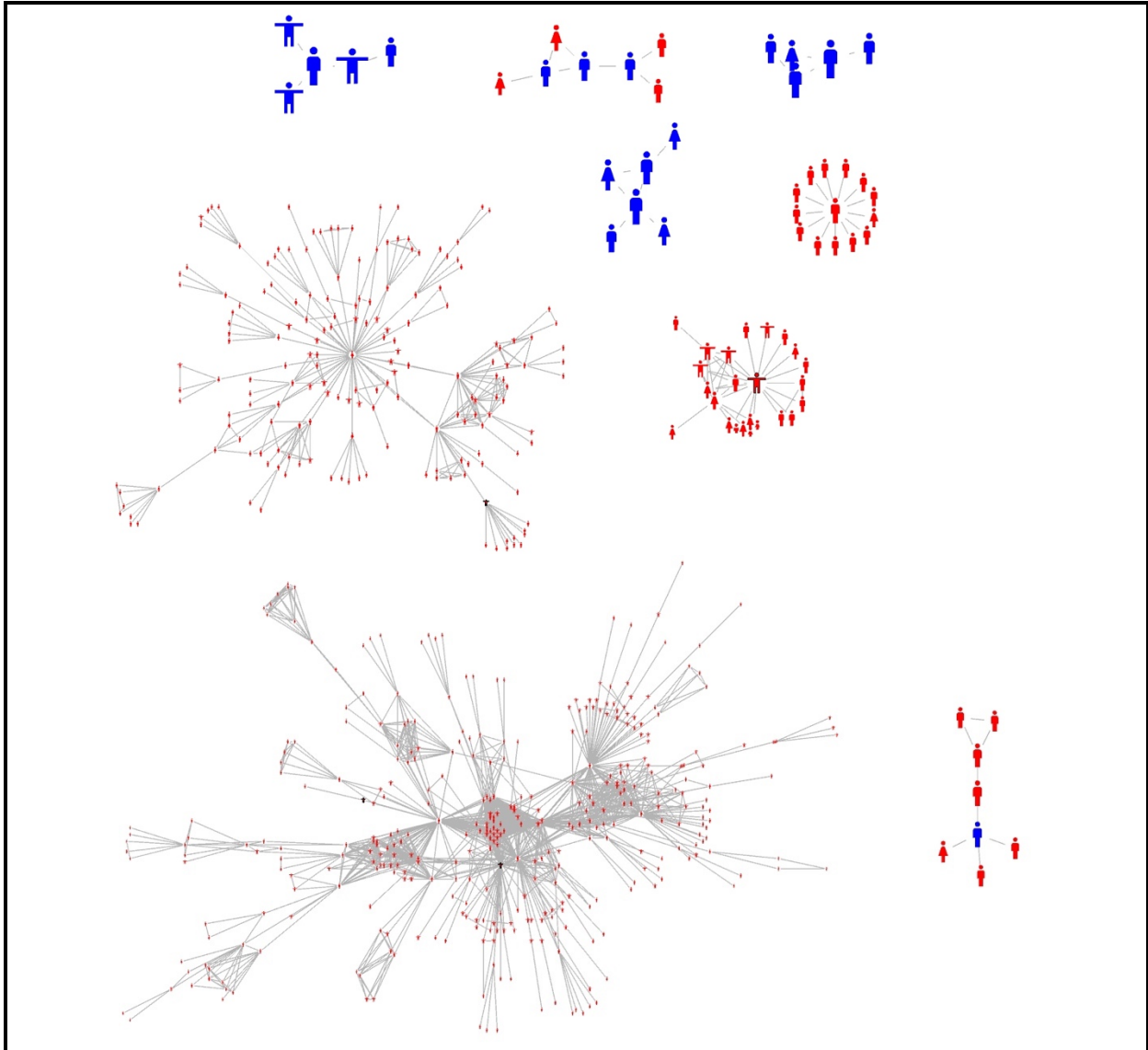


Figure 82 Pondération de la taille des nœuds par leur centralité de proximité (non standardisée).
Les nœuds bleus sont les 20 acteurs ayant les plus fortes proximités.

De même, la centralité de proximité standardisée, représentant l'inverse de la moyenne des distances entre un nœud v_i et tous les autres nœuds de sa composante, favorisera les nœuds ayant de courtes distances moyennes avec les autres acteurs de leurs composantes respectives, c'est-à-dire ceux des composantes de faible taille⁴³³ (figure 83).

⁴³³ La centralité de proximité standardisée pourrait cependant être pertinente dans le cas d'un graphe déconnecté dont les composantes connexes sont de taille homogène.

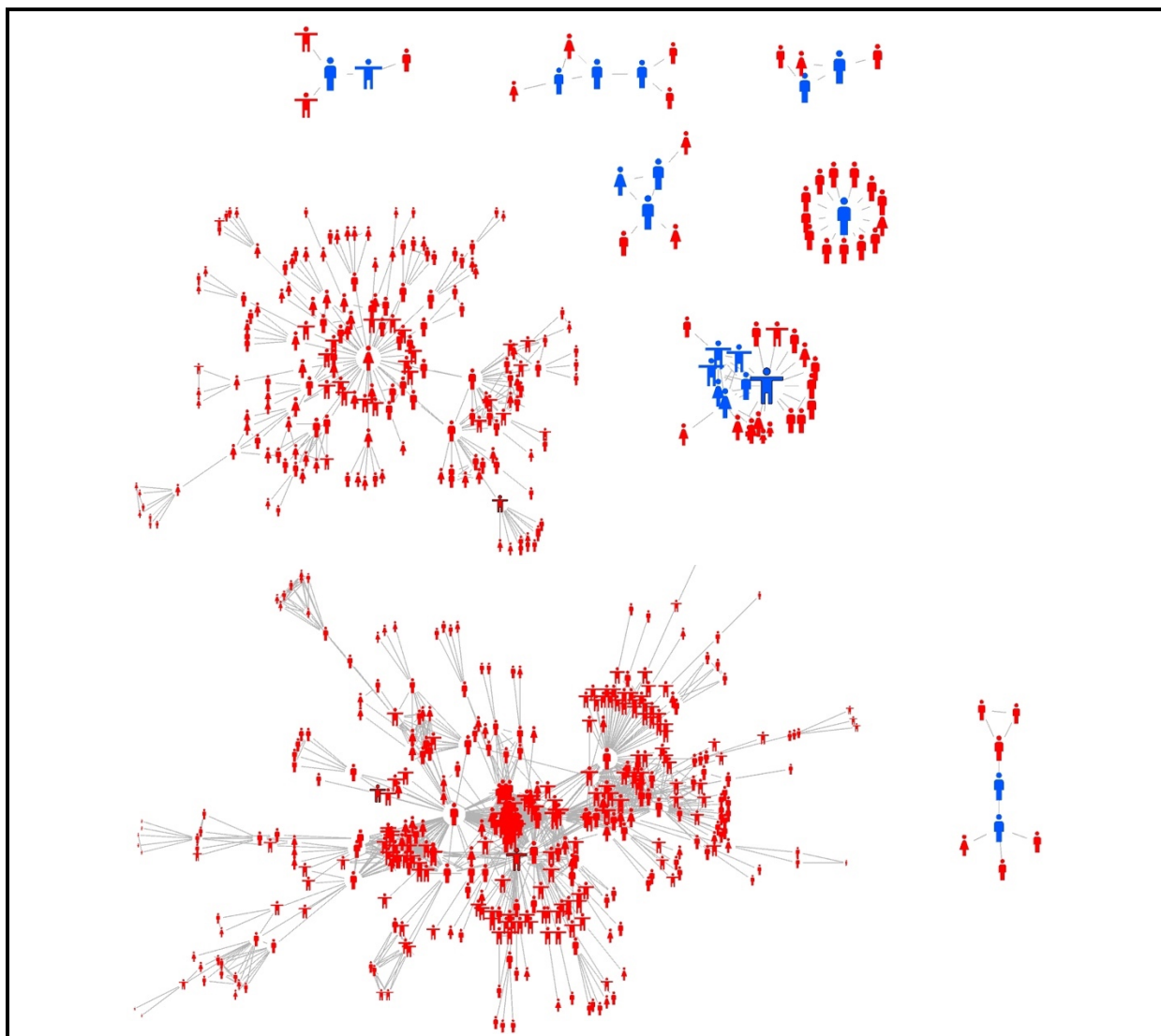


Figure 83 Pondération de la taille des nœuds par leur centralité de proximité standardisée. Les nœuds bleus sont les 20 acteurs ayant les plus fortes proximités.

Cette mesure ne sera donc finalement pas utilisée dans la suite du travail.

c) Centralité d'intermédiation

La centralité d'intermédiation⁴³⁴, qui est également une mesure globale, quantifie le niveau d'intermédiation d'un nœud v_i , c'est-à-dire le fait qu'il soit ou non un passage pour relier plusieurs nœuds du graphe⁴³⁵. Plus il existe de chemins géodésiques entre deux nœuds $(v_j, v_k) \in V$ du graphe passant par v_i , plus la centralité d'intermédiation de ce dernier est élevée. On parle alors de position stratégique – ou non⁴³⁶, et d'importance dans le réseau.

⁴³⁴ Dans un souci de lisibilité et par abus de langage, nous emploierons également le terme « intermédiation » pour désigner la centralité d'intermédiation.

⁴³⁵ Michel Lacroix, « Littérature, analyse de réseaux et centralité », *op. cit.*

⁴³⁶ Carlo Morselli, « Assessing Vulnerable and Strategic Positions in a Criminal Network », *op. cit.*

La centralité d'intermédiation identifie les « connecteurs » (ou « *brokers* ») dans le réseau, *i.e.* les acteurs permettant de relier *via* des trajets courts deux nœuds ou deux parties distinctes du réseau, parfois en étant un intermédiaire obligatoire⁴³⁷. D'avantage de chemins passent par les acteurs ayant une centralité d'intermédiation élevée, et donc d'avantage d'informations et de ressources, que ces acteurs peuvent ou non laisser passer. En plus de contrôler les flux d'information et/ou de ressources entre plusieurs parties du réseau, ces acteurs bénéficient d'un accès privilégié à de multiples sources, et peuvent agir en intermédiaires. Enfin, leur importance stratégique est non négligeable, dans la mesure où sans ces individus, certaines parties du réseau pourraient se trouver complètement déconnectées les unes des autres. La centralité d'intermédiation du nœud v_i correspond à la somme des rapports, pour chaque paire de nœuds $(v_j, v_k) \in V$, du nombre chemins géodésiques entre v_j et v_k passant par v_i sur le nombre total de chemins géodésiques entre v_j et v_k .

Soient $G = (V, E)$ un graphe d'ordre N , σ_{jk} le nombre total de chemins géodésiques entre les nœuds v_j et v_k , $\sigma_{jk}(v_i)$ le nombre de chemins géodésiques entre les nœuds v_j et v_k (avec $j, k \neq i$) passant par v_i . La centralité d'intermédiation $C_{intermédiation}(v_i)$ du nœud $v_i \in V$ est définie comme suit⁴³⁸ :

$$C_{intermédiation}(v_i) = \sum_{j \in V} \sum_{k > j \in V} \frac{\sigma_{jk}(v_i)}{\sigma_{jk}}$$

La valeur maximale de la centralité d'intermédiation est obtenue pour un nœud v_i appartenant à tous les chemins géodésiques entre toutes les paires de nœuds $(v_j, v_k) \in V$ (où $j, k \neq i$ et $k > j$) du graphe. Cette valeur maximale est égale à $\frac{(N-1) \cdot (N-2)}{2}$ ⁴³⁹. La centralité d'intermédiation est ainsi également dépendante de la taille du graphe. Comme la centralité de degré, elle peut être standardisée sur l'effectif total N du graphe (qu'il soit connecté ou non). Afin de standardiser cette mesure pour la rendre comparable d'un graphe à l'autre, Stanley Wasserman et Katherine Faust⁴⁴⁰ proposent de mesurer la proportion de chemins

⁴³⁷ Stephen P. Borgatti, Martin G. Everett et Jeffrey C. Johnson, *Analyzing social networks*, *op. cit.*

⁴³⁸ Linton C. Freeman, « Centrality in social networks conceptual clarification », *op. cit.* ; Stanley Wasserman et Katherine Faust, *Social Network Analysis Methods and Applications*, *op. cit.*

⁴³⁹ $(N - 1) \cdot (N - 2)$ étant le nombre de paires de nœuds, et le facteur $\frac{1}{2}$ permettant, dans un graphe non dirigé où (v_j, v_k) équivaut à (v_k, v_j) de ne comptabiliser qu'une seule fois chaque paire de nœuds.

⁴⁴⁰ Stanley Wasserman et Katherine Faust, *Social Network Analysis Methods and Applications*, *op. cit.*

géodésiques auxquels appartient v_i ou, en d'autres termes, le pourcentage de chances que l'information passe par v_i lorsqu'elle circule entre deux nœuds $(v_j, v_k) \in V$ pris au hasard.

$$C'_{\text{intermédierité}}(v_i) = \frac{2}{(N-1) \cdot (N-2)} \sum_{j \in V} \sum_{k > j \in V} \frac{\sigma_{jk}(v_i)}{\sigma_{jk}}$$

La figure 84 représente le graphe *INR* en pondérant la taille des nœuds par leur centralité d'intermédierité.



Figure 84 Pondération de la taille des nœuds par leur centralité d'intermédierité standardisée. Les nœuds orange sont les 20 acteurs ayant les plus fortes intermédierités.

Le tableau 64 reporte, pour ces 20 acteurs, les valeurs des centralités d'intermédierité ainsi que les valeurs standardisées sur l'ensemble du graphe.

Tableau 64 Centralité d'intermédiarité et centralité d'intermédiarité centralisée des 20 premiers nœuds, classés par centralité décroissante

ID_anonym	Betweenness	Betweenness_norm
1	24848,60	13,20%
348	18493,70	9,83%
1115	14208,60	7,55%
250	14012,40	7,45%
110	9944,75	5,28%
2	6990,15	3,71%
345	6479,28	3,44%
344	6479,28	3,44%
402	6069,66	3,23%
9	4150,08	2,21%
15	3836,46	2,04%
45	3553,33	1,89%
27	3439,80	1,83%
11	3397,67	1,81%
16	3241,50	1,72%
115	2688,50	1,43%
23	2396,98	1,27%
379	2383,56	1,27%
243	2222,09	1,18%
48	2199,15	1,17%

Le nœud le plus intermédiaire est ainsi compris dans 13,20% des chemins géodésiques du graphe. Sur un total de 615 acteurs, 379 (soit 62%) ont une intermédiarité nulle (n'appartiennent à aucun chemin géodésique) ; 208 (soit 33,82%) appartiennent à moins de 1% des chemins géodésiques ; et seulement 26 acteurs (soit 4,22%) appartiennent à 1% ou plus de ces chemins.

La figure 85 illustre la distribution et la répartition des valeurs de la centralité d'intermédiarité (standardisée) pour l'ensemble des acteurs ; le tableau 65 en reporte les indicateurs statistiques de dispersion. Pour mieux en saisir le sens, ces derniers sont, comme les visualisations, donnés pour les valeurs de centralité d'intermédiarité standardisées⁴⁴¹ (c'est-à-dire que nous parlons en termes de pourcentage d'appartenance à l'un des chemins géodésiques existants entre deux nœuds quelconques).

⁴⁴¹ En effet, la valeur absolue de la centralité d'intermédiarité est peu parlante en soi : elle représente le nombre de chemins géodésiques du graphe dont le nœud d'intérêt fait partie, sur l'ensemble des chemins géodésiques existants dans le graphe. Il est ainsi plus tangible de s'exprimer en termes de probabilité d'appartenance à l'un de ces chemins. Pour la même raison, les indicateurs statistiques de moyenne et d'écart-type ne sont ici pas pertinent : nous basons donc cette analyse sur les visualisations graphiques (figure 85) et les indicateurs de dispersion (tableau 65).

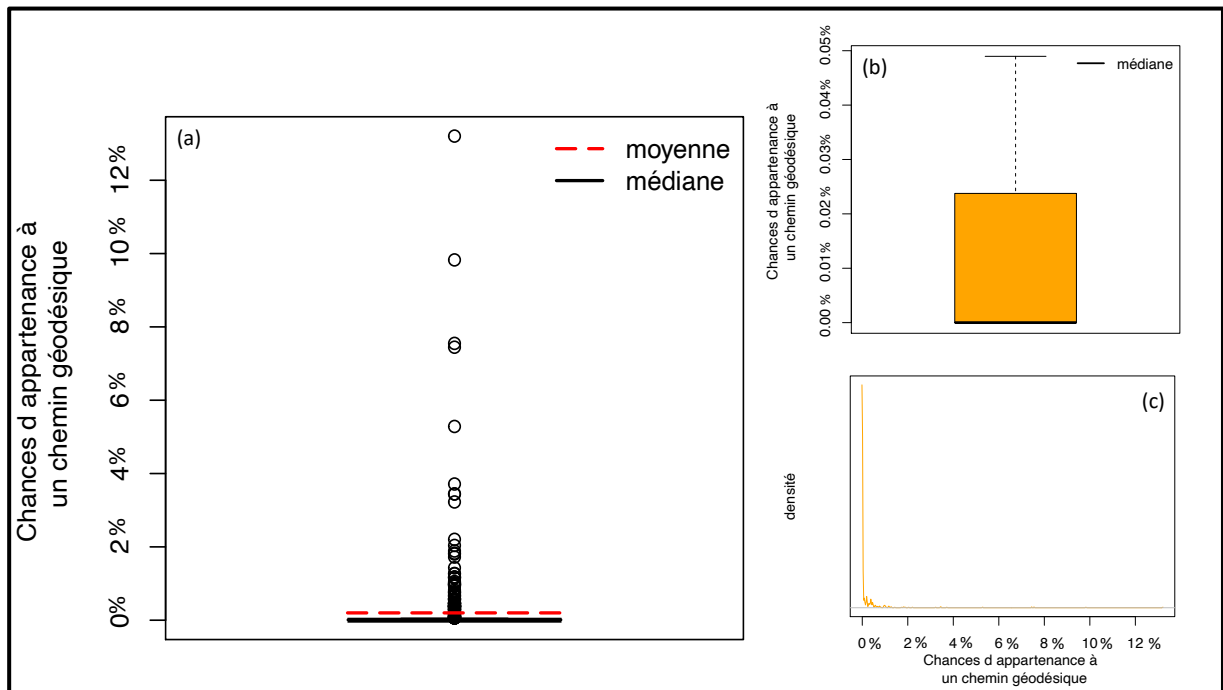


Figure 85 Répartition des valeurs de centralité d'intermédiarité (standardisée). (a) *Boxplot* ; (b) *Boxplot* sans les outliers ; (c) Courbe de densité

Le graphique (a) est un *boxplot* représentant la répartition des valeurs de la centralité d'intermédiarité standardisée. Le graphique (b) est également un *boxplot* des valeurs de la centralité d'intermédiarité standardisée, duquel ont été retirées les valeurs extrêmes afin de mieux visualiser la répartition des valeurs dans leur ensemble. Le graphique (c) représente la densité des chances d'appartenance des acteurs à un chemin géodésique entre deux nœuds quelconques (en pourcentage) sur l'ensemble des 615 acteurs.

Tableau 65 Indicateurs de position et de dispersion pour la centralité d'intermédiarité

Indicateur	Chances	Signification
1 ^{er} quartile <i>Q1</i>	0%	25% des acteurs ne font partie d'aucun chemin géodésique.
Médiane	0%	50% des acteurs ne font partie d'aucun chemin géodésique.
3 ^{ème} quartile <i>Q3</i>	0,02%	75% des acteurs font partie de 0,02% ou moins des chemins géodésiques.
Écart interquartile (<i>IQR = Q3 - Q1</i>)	0,02%	Cet indicateur permet de détecter les valeurs extrêmes. Il s'agit ici des centralités d'intermédiarité des acteurs appartenant à 0,06% ou plus de chemins géodésiques. Cela concerne 23,8% des acteurs.
Min	0%	Les acteurs les moins intermédiaires n'appartiennent à aucun chemin géodésique.
Max	13,2%	L'acteur le plus intermédiaire appartient à 13,2% des chemins géodésiques.

Les visualisations graphiques et les indicateurs de position et de dispersion montrent que les valeurs de centralité d'intermédiarité sont extrêmement concentrées dans les valeurs très basses (figure 85 (c), $Q3$) : 76,2% des acteurs ont une probabilité de moins de 0,06% d'appartenir à un chemin géodésique. Les valeurs extrêmes, qui concernent 23,8% des nœuds (soit 146 acteurs) sont en revanche très dispersées : elles s'étendent d'une probabilité d'appartenance à l'un des chemins géodésiques entre deux nœuds choisis au hasard de 0,06% à 13,2%. Compte tenu de la faible connectivité du réseau, ces valeurs mettent en évidence les acteurs les plus intermédiaires dans le réseau.

VII.3 Centralités secondaires

Les centralités secondaires sont de nouvelles mesures de centralités développées dans le but d'identifier de potentiels rôles et positions importants n'étant pas mis en évidence par les mesures de centralité « primaires⁴⁴² » appliquées jusqu'ici. Ces mesures sont inspirées de la centralité de vecteur propre, qui permet de mesurer le niveau de connexion d'un nœud par rapport aux autres nœuds fortement connectés du graphe⁴⁴³. L'idée est qu'un nœud peut aussi être important s'il est fortement connecté aux autres nœuds centraux du graphe, dans la mesure où il bénéficie de la position de ses voisins directs⁴⁴⁴ (figure 86). La centralité de vecteur propre est une mesure récursive, c'est-à-dire qu'elle dépend, pour le nœud v_i , des centralités de vecteur propre de ses voisins, qui sont elles-mêmes fonction de la centralité de vecteur propre de v_i . Nous avançons cependant que la notion d'importance gagnerait à ne pas être mesurée de cette manière récursive uniquement, et qu'elle est également traduite par les centralités de degré et d'intermédiarité. Nous proposons donc de créer deux nouvelles mesures, appelées centralités secondaires de degré (C_{2deg}) et d'intermédiarité (C_{2inter}), permettant de calculer à quel point un nœud du graphe est connecté aux nœuds ayant respectivement une forte centralité de degré et d'intermédiarité.

L'exemple de la figure 86 illustre une situation où un acteur peut disposer d'avantages non pas grâce à sa propre position, mais à celle de ses voisins directs. La personne D n'est centrale en termes ni de degré, ni d'intermédiarité. Cependant, étant connectée à la personne J

⁴⁴² Le terme « centralité primaire » est employé par opposition aux centralités secondaires. Il fait référence aux centralités évoquées jusqu'ici.

⁴⁴³ Le tableau des centralités de degré de ces 62 acteurs est disponible en annexe 16.

⁴⁴⁴ Ronald S. Burt, *Brokerage and closure: an introduction to social capital*, Oxford, Oxford University Press, 2005, 279 p.

dont ces centralités sont élevées, D pourra bénéficier des contacts, de l'influence et de l'importance que cette position confère à J.

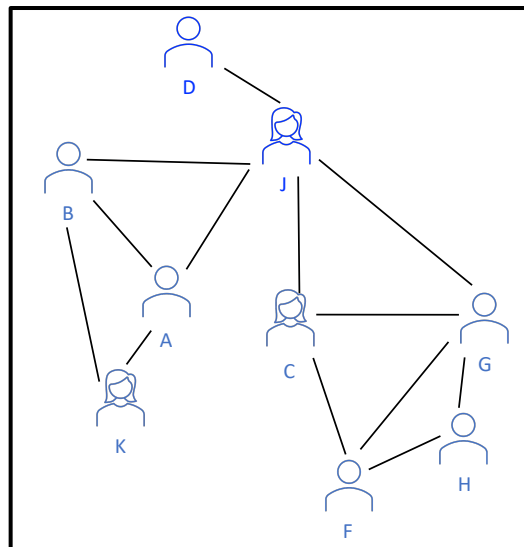


Figure 86 Exemple de position importante non détectée par les mesures de centralité primaires

Les centralités secondaires sont inspirées du calcul de la centralité de vecteur propre, qui se calcule comme suit :

$$\mu \cdot C_{vect}(v_i) = a_{i1} \cdot C_{vect}(v_1) + a_{i2} \cdot C_{vect}(v_2) + \dots + a_{in} \cdot C_{vect}(v_n)$$

Les valeurs des centralités secondaires s'obtiennent pour le nœud v_i en faisant la somme des centralités standardisées de degré (respectivement d'intermédiarité) des nœuds auxquels il est directement connecté.

a) Centralité secondaire de degré

La centralité secondaire de degré⁴⁴⁵ est un indicateur permettant de déterminer à quelle proportion du réseau un acteur peut accéder grâce à ses voisins directs. Pour chaque nœud, sa valeur s'obtient comme suit :

$$C_{2deg}(v_i) = \frac{1}{N-1} \cdot (a_{i1} \cdot (C_{deg}(v_1) - 1) + a_{i2} \cdot (C_{deg}(v_2) - 1) + \dots + a_{in} \cdot (C_{deg}(v_n) - 1))$$

$$C_{2deg}(v_i) = \frac{1}{N-1} \cdot \sum_j a_{ij} \cdot (C_{deg}(v_j) - 1)$$

⁴⁴⁵ Dans un souci de lisibilité et par abus de langage, nous emploierons également le terme « degré secondaire » pour désigner la centralité secondaire de degré.

Afin de rester à des échelles analogues pour par la suite comparer les différents indicateurs de centralité, cette valeur est rapportée au nombre maximal de liens possibles, $N - 1$ ⁴⁴⁶. Nous retirons systématiquement 1 à la centralité de degré des nœuds v_j adjacents à v_i afin de ne pas comptabiliser le lien entre v_i et v_j dans le calcul du degré secondaire de v_i (figure 87). Deux causes peuvent produire une valeur élevée de centralité secondaire de degré : soit car v_i est connecté à un grand nombre de nœuds, dont la somme des centralités finit par atteindre une valeur importante⁴⁴⁷, soit parce que certains de ses voisins ont une centralité élevée. C'est ce deuxième cas qui nous intéresse. La figure 87 illustre le principe de fonctionnement de ce nouvel indicateur.

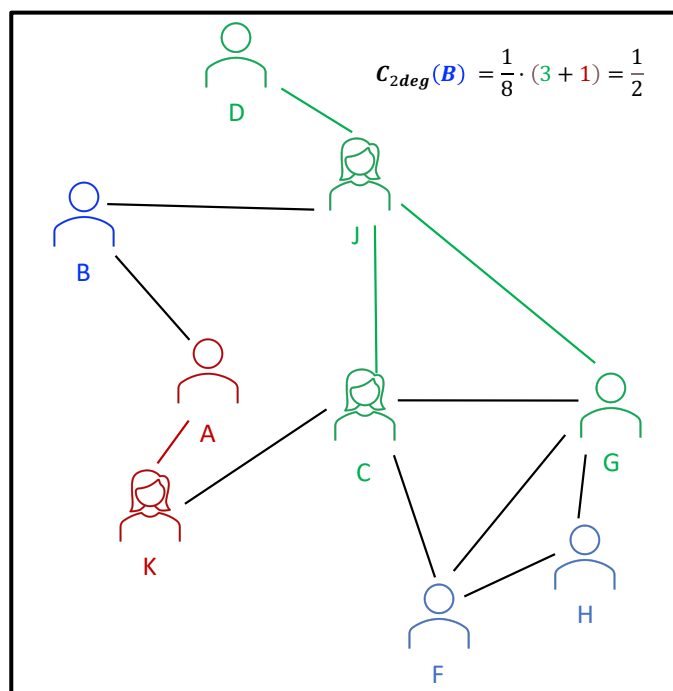


Figure 87 Exemple de faible centralité primaire de degré mais de fort degré secondaire.

Dans cet exemple, le nœud B a un degré de 2, assez faible par rapport aux autres acteurs. Cependant, il a accès à respectivement 3 et 1 acteurs supplémentaires grâce à sa relation avec J et avec A (J et A sont respectivement de degré 4 et 2, mais on retire 1 à leurs degrés respectifs afin de ne pas comptabiliser les arêtes J-B et A-B). La valeur absolue de la centralité secondaire n'a pas de sens tangible en soi. Il ne s'agit pas d'une mesure à proprement parler, mais plutôt d'un indicateur, qui pourra être utilisé comme critère de comparaison entre des acteurs.

⁴⁴⁶ Nous ne parlons pas ici de « standardisation » comme pour les centralités de degré et d'intermédiarité, car ce dernier implique de diviser les valeurs obtenues par leur maximum possible, ce qui n'est pas le cas ici. Nous divisons simplement la valeur obtenue en sommant les centralités des nœuds adjacents au nœud v_i par le nombre maximum de liens possibles dans le graphe, c'est-à-dire la valeur maximale de la centralité primaire de degré.

⁴⁴⁷ C'est-à-dire que v_i a lui-même une centralité « primaire » de degré importante.

La figure 88 représente le graphe *INR* en pondérant la taille des nœuds par leur centralité primaire de degré. Les 20 premiers acteurs les plus connectés sont représentés en bleu. Les nœuds verts représentent les 10 acteurs possédant la plus forte centralité secondaire de degré, après exclusion de ceux faisant déjà partie du groupe bleu.

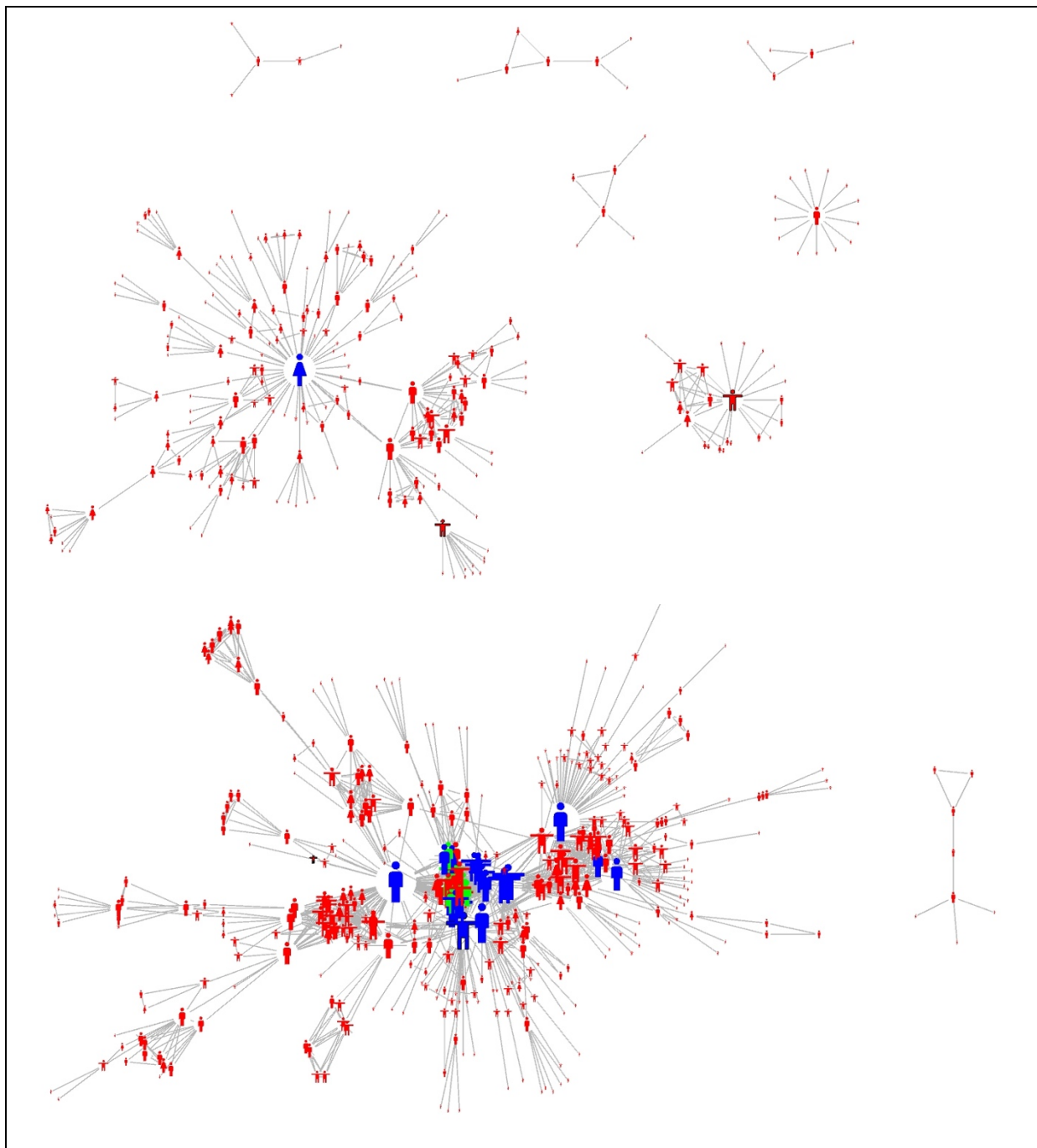


Figure 88 Pondération de la taille des nœuds par leur centralité « primaire » de degré. Les acteurs bleus sont les 20 acteurs ayant les plus forts degrés, préalablement mis en évidence. Les acteurs verts sont les 10 acteurs ayant les plus forts degrés secondaire (après exclusion de ceux appartenant déjà au premier groupe, au nombre de 14).

Les 10 acteurs représentés en vert ne faisant pas partie du groupe bleu, nous émettons l'hypothèse que leur centralité secondaire de degré est élevée car leurs voisins directs ont de forts degrés (et non car ils ont eux-mêmes un fort degré primaire). Ces 10 acteurs appartiennent

tous à la composante S_8 , dont l'effectif est le plus grand ($N_{S_8} = 368$). Étant situés dans la partie la plus dense de S_8 , il est difficile de les distinguer à l'aide de cette visualisation relationnelle.

La figure 89 compare les répartitions de ces deux valeurs.

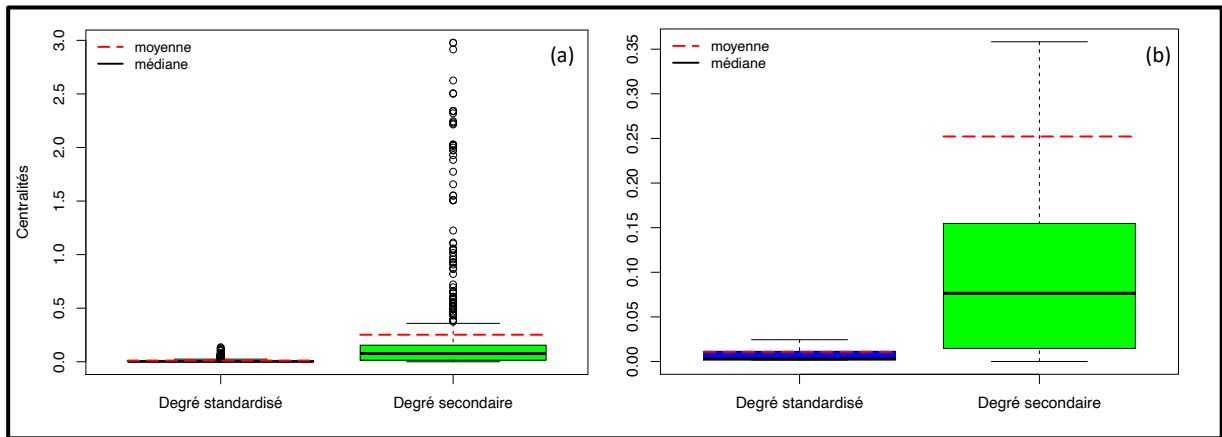


Figure 89 Comparaison des *boxplot* de répartitions des valeurs de degrés primaires et secondaires, avec les valeurs extrêmes (a) et sans (b).

Naturellement, le degré secondaire étant une somme de degrés primaires, ses valeurs sont plus élevées, et plus étendues⁴⁴⁸. Afin de déterminer si cet indicateur peut permettre de détecter des acteurs structurellement ou fonctionnellement importants pour le réseau que la centralité primaire de degré ne permet pas de mettre évidence, nous étudions la corrélation entre ces deux indicateurs. La figure 90 montre les valeurs des degrés secondaires des 615 acteurs en fonction de leurs degrés primaires respectifs. En effet, si ces deux grandeurs sont très fortement corrélées, alors la centralité secondaire n'ajoute que peu d'informations par rapport à celles déjà obtenues grâce à la centralité primaire de degré.

⁴⁴⁸ Raison pour laquelle la figure 88 illustre le réseau avec la taille des nœuds pondérées par la centralité primaire (et non secondaire) de degré, sans quoi la visualisation serait difficilement lisible.

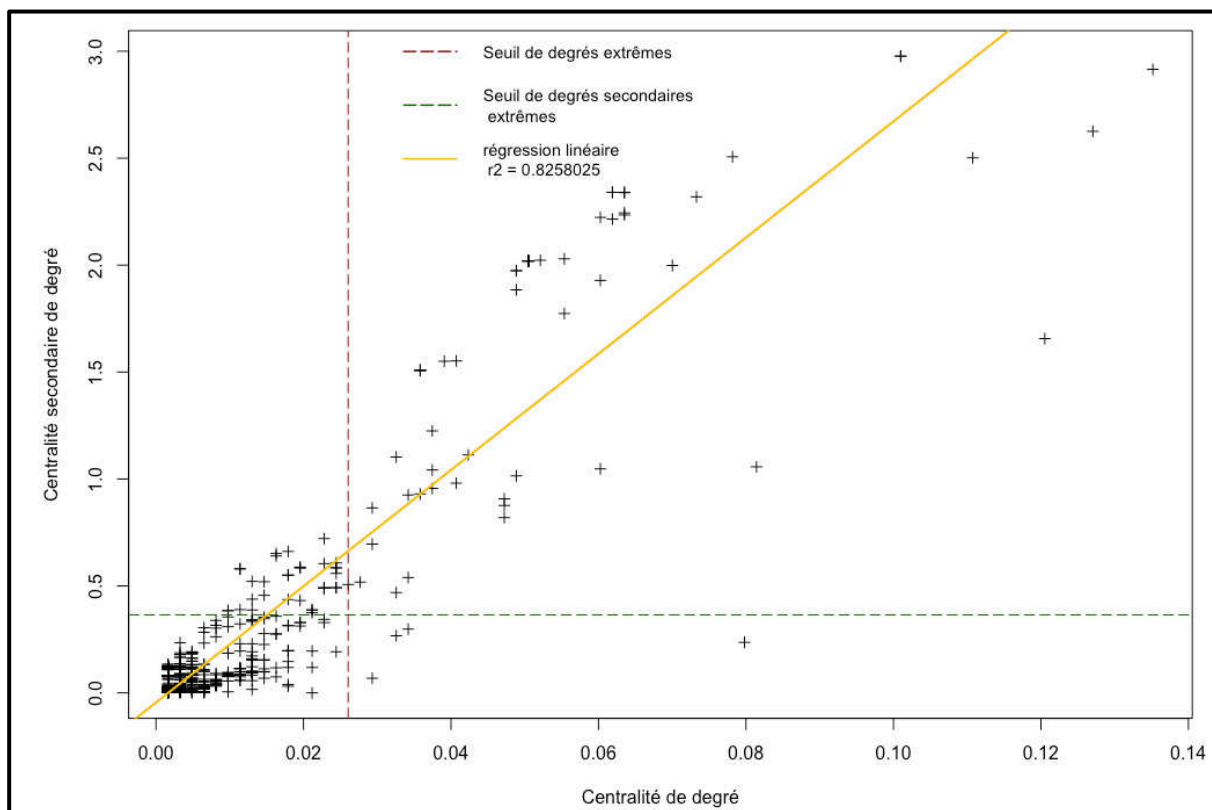


Figure 90 Scatterplot représentant, pour chaque acteur, ses centralités primaires et secondaires de degré. L'équation de régression est : $y = 27.16924 \cdot x - 0.04498$.

Une régression linéaire a été appliquée. Il s'agit d'un modèle cherchant à établir une relation linéaire⁴⁴⁹ entre deux variables x et y décrivant le mieux possible leur relation « réelle »⁴⁵⁰. Ainsi, plus ces deux variables sont corrélées, plus les points du *scatterplot* seront proches de la droite de régression obtenue par le modèle. Le coefficient de corrélation linéaire, r_2 , permet de qualifier la corrélation entre les deux variables. Une corrélation positive⁴⁵¹ parfaite correspondant à $r_2 = 1$, plus ce coefficient est proche de 1, mieux l'équation de régression décrit l'évolution de y en fonction x ⁴⁵². Le calcul de r_2 se fait par la méthode dite « des moindres carrés » (figure 91), qui consiste à déterminer l'équation de régression de telle sorte que la somme des carrés des distances (d_i) entre chaque point du graphe (en noir) et sa projection (en vert) sur la droite de régression soit minimale. Cette méthode minimise la somme :

$$\sum_i d_i^2$$

⁴⁴⁹ De la forme $y = a \cdot x + b$.

⁴⁵⁰ *i.e.* l'évolution de y en fonction de x dans les données.

⁴⁵¹ Une corrélation est positive lorsque les deux grandeurs évoluent dans le même sens. Si y diminue lorsque x augmente (ou inversement), on parle de corrélation négative. Pour une corrélation négative parfaite, $r_2 = -1$. Enfin, pour une corrélation inexistante, le coefficient r_2 vaut 0.

⁴⁵² Jean-Marie Bouroche et Gilbert Saporta, *L'analyse des données*, Paris, Presses Universitaires de France, 2005, 128 p.

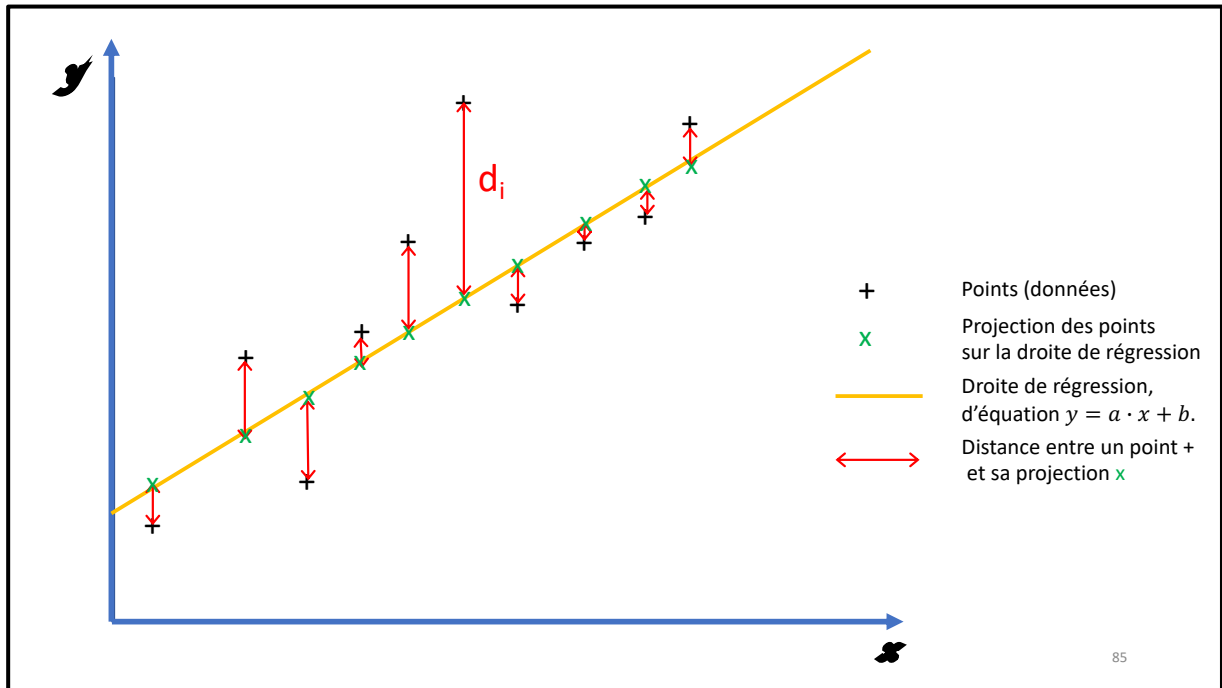


Figure 91 Schéma de la méthode des moindres carrés

Dans notre cas, x correspond au degré primaire, y au degré secondaire. La corrélation est positive : les deux indicateurs évoluent dans le même sens. Le coefficient de régression obtenu est : $r^2 = 0.8258$.

Il y a donc une forte corrélation entre ces valeurs, ce qui signifie que, globalement, c'est-à-dire à l'échelle du réseau, la centralité secondaire de degré (telle que nous l'avons définie) ne permet pas particulièrement de détecter de nouveaux acteurs potentiellement importants par rapport à ceux déjà mis en évidence par la centralité primaire de degré. Cependant, à l'échelle individuelle, cet indicateur pourrait aider à comprendre le rôle et la position d'un acteur donné dans l'organisation criminelle, notamment pour les points les plus éloignés de la droite de régression (figure 90). Cette hypothèse sera discutée dans la section VII.4.

b) Centralité secondaire d'intermédiation

La centralité secondaire d'intermédiation⁴⁵³ est un indicateur permettant de déterminer à quelle proportion des flux circulant dans le réseau un acteur peut accéder grâce à ses voisins directs. Pour chaque nœud, sa valeur s'obtient par le calcul ci-dessous.

$$C_{2inter}(v_i) = (a_{i1} \cdot C'_{inter}(v_1) + a_{i2} \cdot C'_{inter}(v_2) + \dots + a_{in} \cdot C'_{inter}(v_n))$$
$$C_{2inter}(v_i) = \frac{2}{(N-1) \cdot (N-2)} \sum_j a_{ij} \cdot C_{inter}(v_j)$$

De même que pour la centralité secondaire de degré, la centralité secondaire d'intermédiation du nœud est obtenue en sommant les centralités primaires d'intermédiation de ses voisins directs, et en rapportant cette valeur à la valeur maximale que peut prendre la centralité primaire d'intermédiation.

La figure 92 représente le graphe *INR* en pondérant la taille des nœuds par leur centralité primaire d'intermédiation ; la figure 93 compare les répartitions des valeurs des indicateurs primaire et secondaire d'intermédiation.

⁴⁵³ Dans un souci de lisibilité et par abus de langage, nous emploierons également le terme « intermédiation secondaire » pour désigner la centralité secondaire de d'intermédiation.

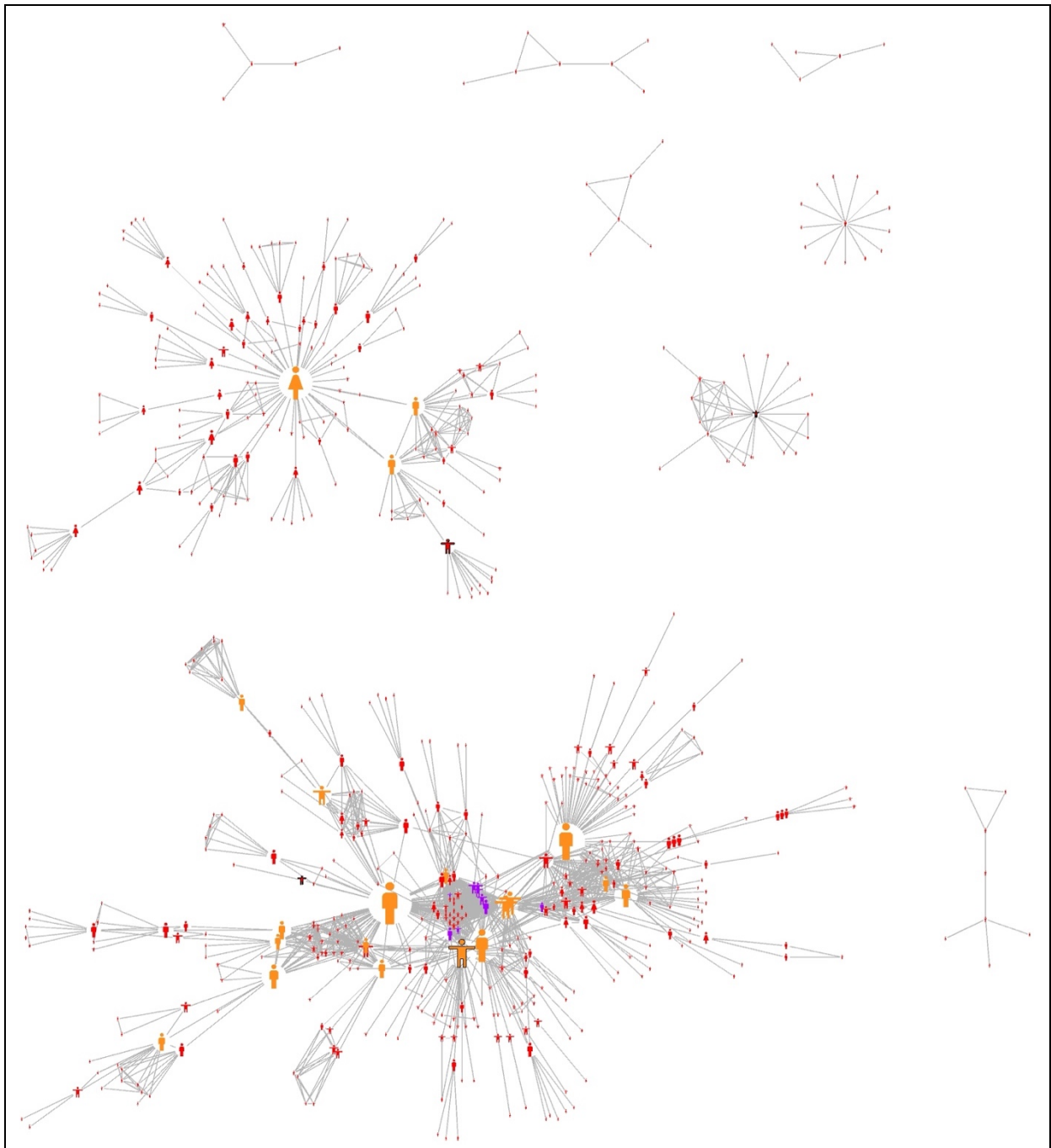


Figure 92 Pondération de la taille des nœuds par leur centralité « primaire » de degré. Les acteurs oranges sont les 20 acteurs ayant les plus fortes intermédialités, préalablement mis en évidence. Les acteurs violets sont les 10 acteurs ayant les plus fortes intermédialités secondaires (après exclusion de ceux appartenant déjà au premier groupe).

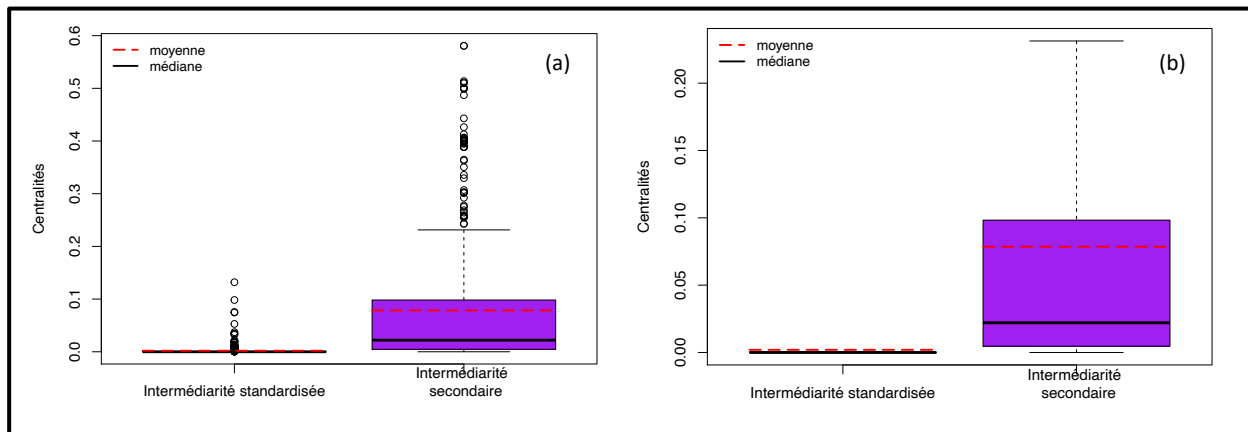


Figure 93 Comparaison des *boxplot* de répartition des valeurs de degrés primaires et secondaires, avec les valeurs extrêmes (a) et sans (b).

A nouveau, la centralité secondaire d'intermédiarité étant une somme d'intermédiarités primaires, ses valeurs sont plus élevées, et plus étendues⁴⁵⁴. Afin d'étudier le potentiel d'information ajoutée de cet indicateur par rapport à celui de la centralité primaire de degré, la figure 94 montre les valeurs des intermédiarités secondaires des 615 en fonction de leurs intermédiarités primaires respectives. En effet, si ces deux grandeurs sont très fortement corrélées, alors la centralité secondaire n'ajoute que peu d'informations par rapport à celles déjà obtenues grâce à la centralité primaire de d'intermédiarité.

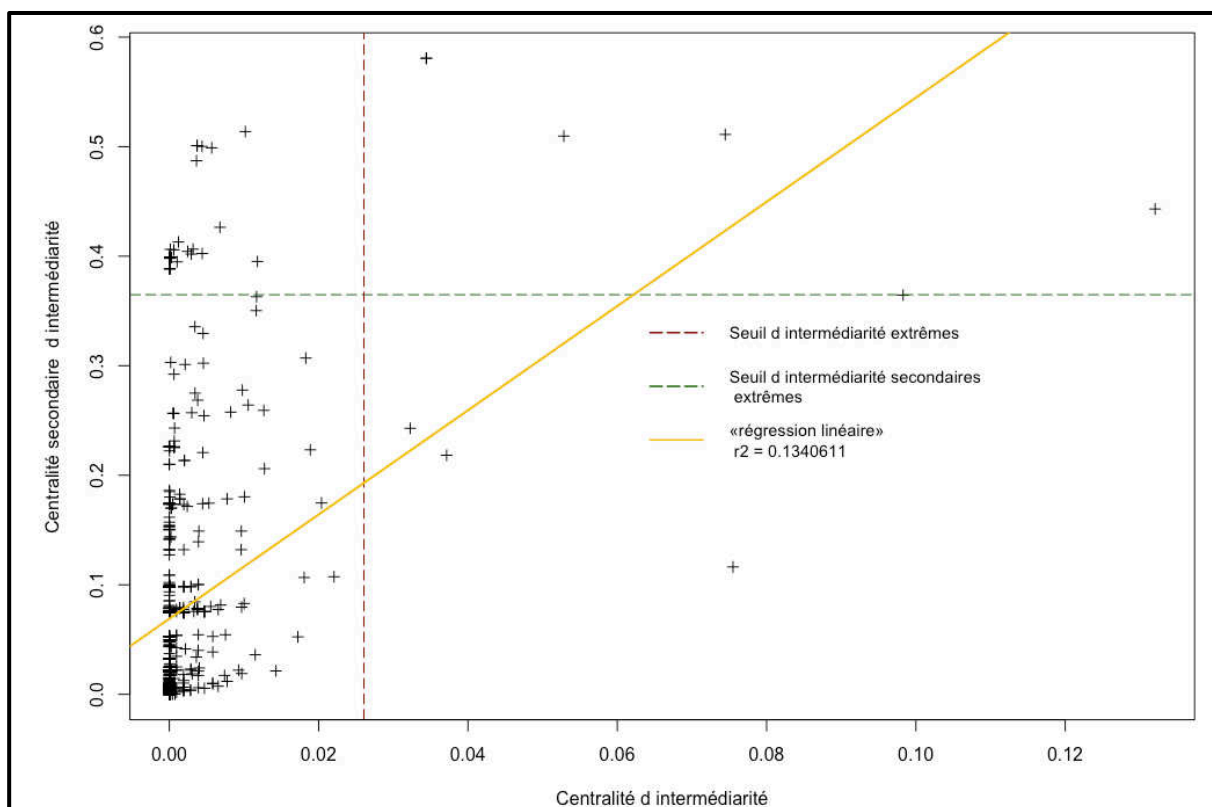


Figure 94 *Scatterplot* représentant, pour chaque acteur, ses centralités primaires et secondaires de degré.

⁴⁵⁴ Raison pour laquelle la figure 92 illustre le réseau avec la taille des nœuds pondérées par la centralité primaire (et non secondaire) de degré, sans quoi la visualisation serait difficilement lisible.

La visualisation graphique montre que la corrélation linéaire n'est pas significative entre ces deux indicateurs : la droite de régression ne s'adapte pas du tout à la répartition des points. Le coefficient de régression est d'ailleurs très faible : $r^2 = 0.1340611$. Nous émettons donc l'hypothèse que la centralité secondaire d'intermédiarité peut permettre de détecter de nouveaux acteurs importants.

VII.4 Détection des positions stratégiques

De précédentes recherches se sont intéressées à l'adaptation de l'interprétation des différentes mesures de centralité dans le contexte spécifique de réseaux criminels. Dans un réseau social (non criminel), un fort degré de centralité est souvent un indicateur d'influence ou de contrôle. Dans le cadre d'un réseau criminel, selon Malcom K. Sparrow, un fort degré de centralité dénote une position de force⁴⁵⁵, dans le sens où le retrait de l'acteur affaiblirait le réseau ; en revanche, selon Marilyn B. Peterson⁴⁵⁶, cela implique une forte visibilité dans le réseau, et donc une certaine vulnérabilité, car l'acteur est alors facilement détectable par les forces de l'ordre ou même par les membres d'organisations concurrentes. Dans cette lignée, les résultats de Wayne E. Baker and Robert R. Faulkner ont montré que les acteurs avec les plus grands degrés de centralité étaient plus souvent condamnés, avec des peines plus lourdes⁴⁵⁷. Ils suggèrent ainsi dans leur étude que les acteurs les plus stratégiques préfèrent demeurer en périphérie de l'organisation en limitant leur nombre d'interlocuteurs pour minimiser les risques de détection par les forces de l'ordre.

Carlo Morselli, dans une étude sur le positionnement des acteurs au sein d'un réseau de distribution de stupéfiants entourant une organisation criminelle bien connue, a montré une forte corrélation entre la probabilité d'arrestation et la centralité de degré. En revanche, cette étude n'a pas mis en évidence de corrélation significative⁴⁵⁸ entre la centralité d'intermédiarité et la probabilité d'arrestation⁴⁵⁹. Pour Ronald S. Burt, la centralité d'intermédiarité est ainsi bien plus pertinente que la centralité de degré pour repérer les acteurs les plus stratégiques : les individus ayant une forte centralité d'intermédiarité, appelés « connecteurs » (ou « *brokers* ») adoptent une approche stratégique et rationnelle (en termes de gain/effort/risque⁴⁶⁰) et restent

⁴⁵⁵ Malcolm K. Sparrow, « The application of network analysis to criminal intelligence », *op. cit.*

⁴⁵⁶ Marilyn B. Peterson, *Applications in criminal analysis: a sourcebook*, Westport, Conn, Greenwood Press, 1994, 312 p.

⁴⁵⁷ Wayne E. Baker et Robert R. Faulkner, « The Social Organization of Conspiracy: Illegal Networks in the Heavy Electrical Equipment Industry », *American Sociological Review*, vol. 58, n° 6, 1993, pp. 837-860.

⁴⁵⁸ La valeur de l'indicateur de corrélation était positive (ce qui signifie que les grandeurs évoluent dans le même sens), mais très faible.

⁴⁵⁹ Carlo Morselli, « Assessing Vulnerable and Strategic Positions in a Criminal Network », *op. cit.*

⁴⁶⁰ Cf. section II.1 b) « Développement de la sociologie de la déviance ».

en retrait, tout en se positionnant comme intermédiaires essentiels. En effet, les acteurs situés de part et d'autre du « pont social » que crée un connecteur dépendent de ce dernier pour accéder aux ressources et aux informations échangées dans le réseau, le connecteur en contrôlant – plus ou moins partiellement – la circulation (figure 95).

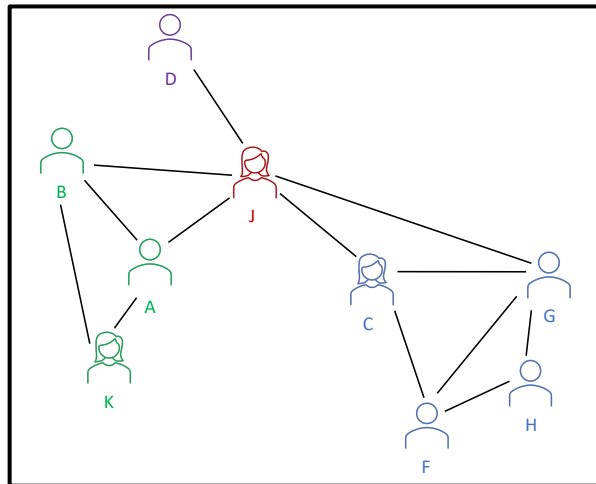


Figure 95 Exemple de *broker*. L'acteur J est ici un intermédiaire essentiel entre les groupes bleu, vert et violet du réseau. Les différents acteurs dépendent donc de J pour accéder aux ressources et informations des groupes dont ils ne font pas partie.

Finalement, selon Carlo Morselli, la compréhension du positionnement structurel des acteurs dans les réseaux criminels nécessite une évaluation et une prise en compte de ces deux formes de centralité combinées⁴⁶¹. La figure 96 situe les 615 acteurs du réseau selon leurs centralités de degré et d'intermédiarité.

⁴⁶¹ Carlo Morselli, « Assessing Vulnerable and Strategic Positions in a Criminal Network », *op. cit.*

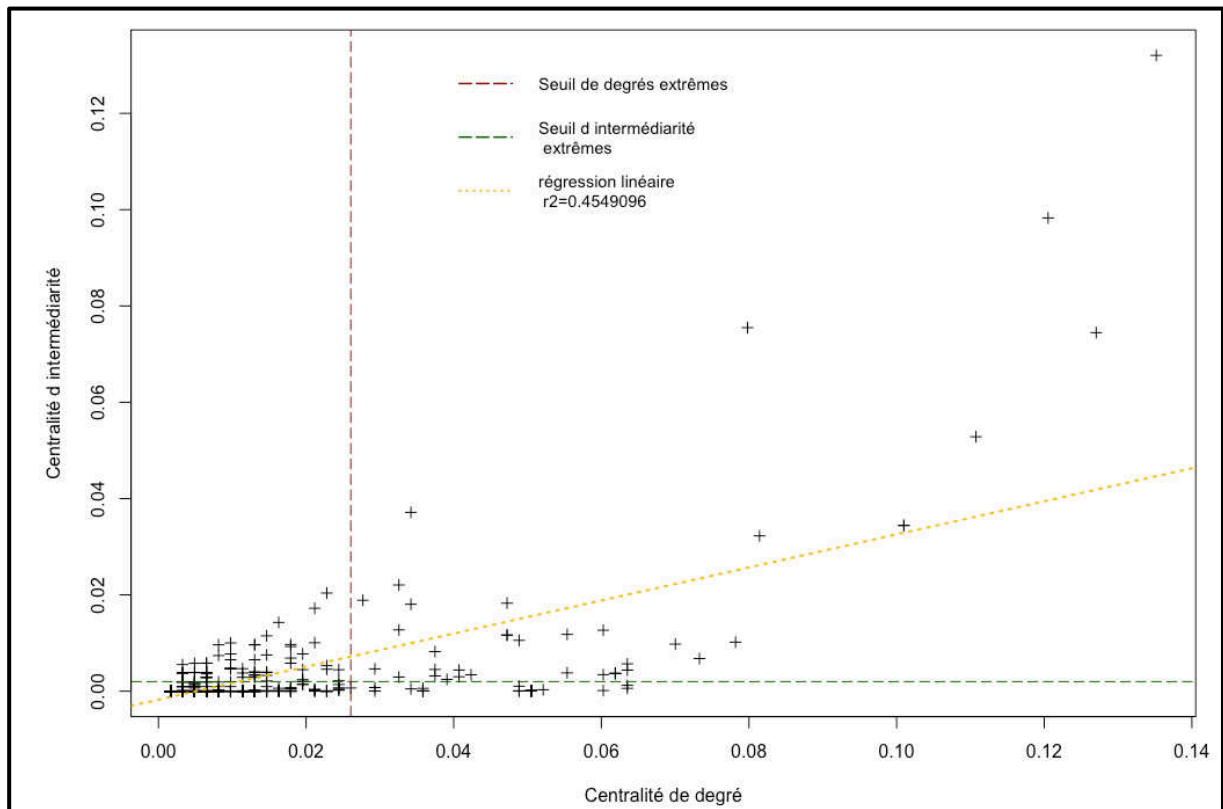


Figure 96 Scatterplot représentant les positions des acteurs dans l'espace graphique des centralités de degré et d'intermédiarité⁴⁶²

Les pointillés représentent la limite au-delà de laquelle les valeurs sont statistiquement considérées comme extrême par rapport à leur distribution. Nous désignons dans la suite de cette section les valeurs supérieures (respectivement inférieures) à cette limite comme fortes (respectivement faibles). La valeur du r^2 indiquant que la corrélation entre ces deux indicateurs n'est pas significative – ce qui rejoint les résultats de Carlo Morselli évoqués précédemment dans cette section⁴⁶³ – nous étudions leur combinaison. Sur les bases des limites de valeurs extrêmes, quatre profils spatiaux de centralités sont ainsi distingués (tableau 66) dans ce que nous appelons l'espace graphique des centralités de degré et d'intermédiarité.

Tableau 66 Profils de centralités dans l'espace graphique degré / intermédiarité

Profil	Caractéristiques	Effectifs
1	Faible degré Faible intermédiarité	450
2	Fort degré Faible intermédiarité	19
3	Fort degré Forte intermédiarité	44
4	Faible degré Forte intermédiarité	102

⁴⁶² Afin de travailler à des échelles comparables, les valeurs standardisées (bornées entre [0 ; 1]) ont été utilisées.

⁴⁶³ Carlo Morselli, « Assessing Vulnerable and Strategic Positions in a Criminal Network », *op. cit.*

Compte tenu du grand nombre d'individus du profil 1, qui rendrait le graphique illisible, tous les points ne sont pas visibles sur la figure 96. Cela n'est pas gênant pour l'analyse, dans la mesure où nous nous intéressons en particulier aux profils 3 et 4 : d'une part, les acteurs possédant un fort degré et une forte intermédiation (au nombre de 44), d'autre part ceux possédant un faible degré, mais une forte intermédiation (au nombre de 102). La liste des acteurs⁴⁶⁴ dans chacun de ces profils est disponible en annexe 17.

Nous nous intéressons également à la position des acteurs selon leurs centralités primaire de degré et secondaire d'intermédiation (figure 97).

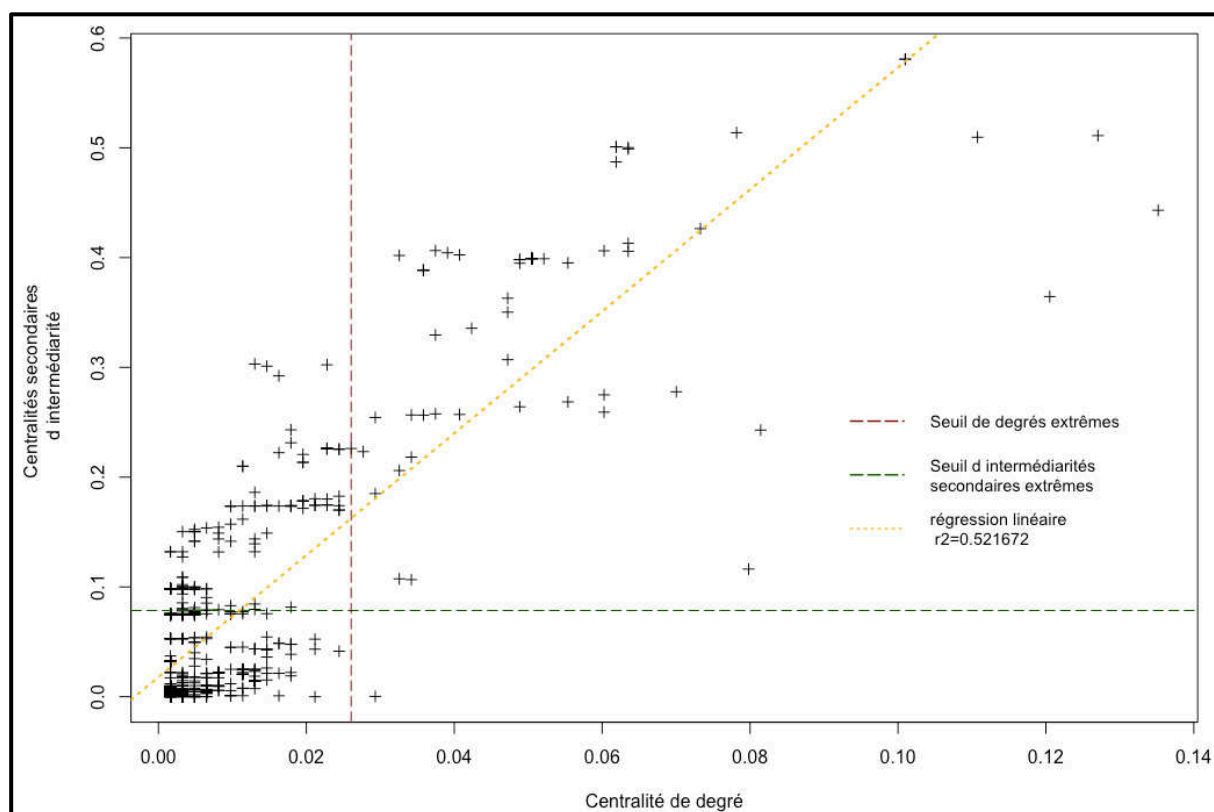


Figure 97 Scatterplot représentant les positions des acteurs dans l'espace graphique des centralités primaire de degré et secondaire d'intermédiation.

La valeur du r^2 indiquant que la corrélation entre ces deux indicateurs n'est pas significative, nous étudions leur combinaison (tableau 67), en particulier les profils 7 (54 acteurs) et 8 (5 acteurs). La liste des acteurs⁴⁶⁵ de chacun de ces profils est disponible en annexe 17.

⁴⁶⁴ Désignés par leur identifiant d'anonymisation, conformément à la circulaire ministérielle n°30053 du 3 septembre 2019.

⁴⁶⁵ *Idem*

Tableau 67 Profils de centralités dans l'espace graphique degré primaire / intermédiation secondaire

Profil	Caractéristiques	Effectifs
5	Faible degré Faible intermédiation secondaire	547
6	Fort degré Faible intermédiation secondaire	9
7	Fort degré Forte intermédiation secondaire	54
8	Faible degré Forte intermédiation secondaire	5

L'intégralité de la démarche analytique a été effectuée à l'aveugle. Cela signifie que, bien que nous disposions de ces informations depuis juin 2020, les résumés et conclusions de l'enquête nous sont restés jusqu'ici inconnus, de même que les identités des 36 prévenus mentionnés dans le réquisitoire définitif du xx mars 2019 et l'arrêté du xx avril 2019⁴⁶⁶.

Le moment est presque venu de prendre connaissance de ces éléments et de les comparer avec les résultats de nos analyses.

Les résultats des études évoquées précédemment sur les centralités de degré et d'intermédiation⁴⁶⁷ indiquent en effet qu'un fort degré rend les acteurs visibles par les forces de l'ordre, tandis que les acteurs les plus stratégiques ont tendance à se placer comme intermédiaires entre plusieurs parties du réseau, tout en restant assez périphériques. Notamment, dans son étude de 2010 sur le positionnement des acteurs dans un vaste trafic de stupéfiants⁴⁶⁸, Carlo Morselli a démontré que pour les acteurs possédant de fortes centralités de degré et d'intermédiation, la probabilité d'arrestation consécutive à un degré élevé outrepassait les avantages stratégiques conférés par la centralité d'intermédiation.

Ainsi, parmi les quatre profils de centralités auxquels nous nous sommes intéressés, nous distinguons d'une part les profils 3 et 7, caractérisés par des individus ayant à la fois un

⁴⁶⁶ Les dates exactes ne sont volontairement pas données, conformément à la circulaire ministérielle n°30053 du 3 septembre 2019.

⁴⁶⁷ Malcolm K. Sparrow, « The application of network analysis to criminal intelligence », *op. cit.* ; Marilyn B. Peterson, *Applications in criminal analysis*, *op. cit.* ; Wayne E. Baker et Robert R. Faulkner, « The Social Organization of Conspiracy », *op. cit.* ; Ronald S. Burt, *Structural holes*, *op. cit.* ; Ronald S. Burt, *Brokerage and closure*, *op. cit.* ; Carlo Morselli, « Assessing Vulnerable and Strategic Positions in a Criminal Network », *op. cit.*

⁴⁶⁸ Carlo Morselli, « Assessing Vulnerable and Strategic Positions in a Criminal Network », *op. cit.* Cette étude, souvent citée dans le présent manuscrit (et les travaux de Carlo Morselli dans leur ensemble), a été une source d'inspiration tout au long de la thèse ; et ce, tant en termes de compréhension que de méthodes d'analyse de structures de réseaux criminels.

fort degré et une forte intermédiation, et d'autre part les profils 4 et 8, pour lesquels les acteurs ont une forte centralité d'intermédiation, mais une faible centralité de degré.

Compte tenu de ces éléments, nous émettons l'hypothèse que, lors de la comparaison entre les résultats analytiques et les conclusions de l'enquête, une majorité des 36 prévenus de l'affaire sera retrouvée dans les profils 3 et 7, et une minorité d'entre eux dans les profils 4 et 8. Il ne serait pas étonnant que certains correspondent également aux profils 2 et 6 (fort degré, faible intermédiation), que nous ne traiterons pas, dans la mesure où l'objectif est de proposer une nouvelle approche pour la détection des individus structurellement ou fonctionnellement importants dans un réseau criminel⁴⁶⁹. Concernant les profils 4 et 8, l'attendu n'est pas de valider les résultats analytiques en démontrant que l'enquête a effectivement révélé l'importance de ces individus, mais bien de proposer une nouvelle approche pouvant éventuellement guider les investigations futures sur des réseaux criminels. Les notions de fonction et de position commencent ici à se distinguer l'une de l'autre. Une première vision, non exhaustive, consiste à considérer la fonction du point de vue des actes commis par les acteurs, correspondant aux éléments d'une infraction au code pénal, établis sur la base d'éléments probants. Il peut par exemple s'agir d'actes de collecte ou de transfert de fonds issus de transactions illicites, (blanchiment d'argent, article 324-1 du code pénal), de production ou de vente de produits stupéfiants (trafic de stupéfiants, articles 222-34 à 222-39 du code pénal), d'effraction ou de cambriolage de propriété privée (vol aggravé, vol avec effraction, articles 311-4 et 311-5 du code pénal) ... La position peut alors être envisagée non pas du point de vue juridique, mais depuis les perspectives du renseignement criminel, de la compréhension et de la capitalisation de connaissances sur les mécanismes structuraux à l'œuvre dans les réseaux criminels. La position ne se définit donc pas à partir d'éléments infractionnels, mais en termes d'espace social – en l'occurrence criminel. La différenciation de l'espace social en champs sociaux se fait sur la base du développement de différents marchés (et flux) d'offre et de demande de capitaux sociaux, économiques, culturels et symboliques au sens de Pierre Bourdieu⁴⁷⁰.

⁴⁶⁹ En effet, l'étude de la simple centralité de degré n'apporterait, en pratique, pas d'éléments méthodologiques nouveaux aux analystes en renseignement ni aux analystes criminels, qui sont très capables de détecter les individus possédant de forts degrés.

⁴⁷⁰ Gisèle Sapiro, François Denord, Julien Duval, Mathieu Hauchecorne, Johan Heilbron, Franck Poupeau et Hélène Seiler (dir.), *Dictionnaire international Bourdieu*, *op. cit.*

Après lecture du résumé de l'enquête, proposé rétroactivement dans la section III.2 a) « Résumé de l'enquête », nous émettons l'hypothèse que les individus identifiés comme collecteurs de fonds, ainsi que les *sarafs* parisiens, se trouvent dans le profil 3 et/ou 7. En effet, les premiers rencontrant de nombreux clients du réseau pour collecter l'argent liquide et les seconds recevant de très nombreux clients dans leur épicerie pour dépôts et retraits de sommes d'argent en espèces plus ou moins conséquentes, ils devraient avoir un degré important. En outre, leurs fonctions respectives supposent une forte intermédiation entre les clients et les donneurs d'ordre du réseau. Les *sarafs* belges coordonnaient les trajets des collecteurs en France, en Belgique et en Espagne et recevaient directement les ordres de virement émis par les *sarafs* pilotant depuis le Maroc les activités de collecte et de blanchiment de fonds du réseau. Rien n'indique dans les conclusions de l'enquête que les *sarafs* belges aient été en contact direct avec les clients, qu'il s'agisse des particuliers, d'individus auprès de qui les collecteurs récupéraient l'argent liquide ou des commanditaires⁴⁷¹ de ces transferts. Nous émettons donc l'hypothèse que ces acteurs se trouvent dans les profils 4 et/ou 8. Ce pourrait aussi être le cas pour les *sarafs* marocains, qui communiquaient avec les investisseurs aux Pays-Bas et aux Émirats Arabes Unis, ainsi qu'avec des acteurs appartenant aux trafics de stupéfiants qui utilisaient ces services.

VII.5 Comparaison des résultats analytiques et des conclusions de l'enquête judiciaire

Lors de la prise de connaissance, le 10 novembre 2022, des identités des 36 personnes accusées dans le cadre de l'affaire, une colonne supplémentaire a été créée dans le fichier relationnel anonymisé correspondant au sociogramme *INR*⁴⁷². La valeur 1 a été renseignée pour les lignes correspondants aux prévenus, la valeur 0 a été renseignée pour les autres acteurs. Un premier constat déroutant est fait à ce moment : sur les 36 personnes accusées, seules 25 sont présentes parmi les 615 acteurs. Est alors envisagé le fait que malheureusement, la méthodologie d'extraction et de filtrage du réseau social de l'investigation proposée dans les chapitres V et VI a isolé ces personnes. En effet, si les liens induits par le modèle sur la base des chemins reliant les 11 personnes manquantes aux autres acteurs ont une pondération trop faible pour avoir été conservés lors du filtrage du sociogramme *Social_network*, alors ces

⁴⁷¹ *i.e.* les trafiquants de stupéfiants.

⁴⁷² Ce fichier relationnel se présente sous forme d'*edge list*, comportant les identifiants d'anonymisation des protagonistes et les liens les unissant, déterminés lors de l'extraction du graphe *INR*.

personnes sont isolées dans *INR*, et ont donc été supprimées lors de l'étape de nettoyage⁴⁷³. Après vérification dans le fichier de données brutes, il s'avère que ces 11 individus n'y ont pas été saisis : ils n'existent pas dans les données. Ce constat est très perturbant, mais à deux semaines seulement du dépôt du manuscrit de la thèse, il n'est plus possible d'y remédier ; nous continuons donc avec les données à disposition.

Les figures 98 et 99 représentent à nouveau les positions des 615 du réseau modélisé par le graphe *INR* dans les espaces graphiques respectifs des centralités de degré et d'intermédiarité, et de degré primaire⁴⁷⁴ et d'intermédiarité secondaire. Les points rouges correspondent aux prévenus⁴⁷⁵. Le tableau 68 reporte la liste des accusés positionnés dans les quatre profils de centralité.

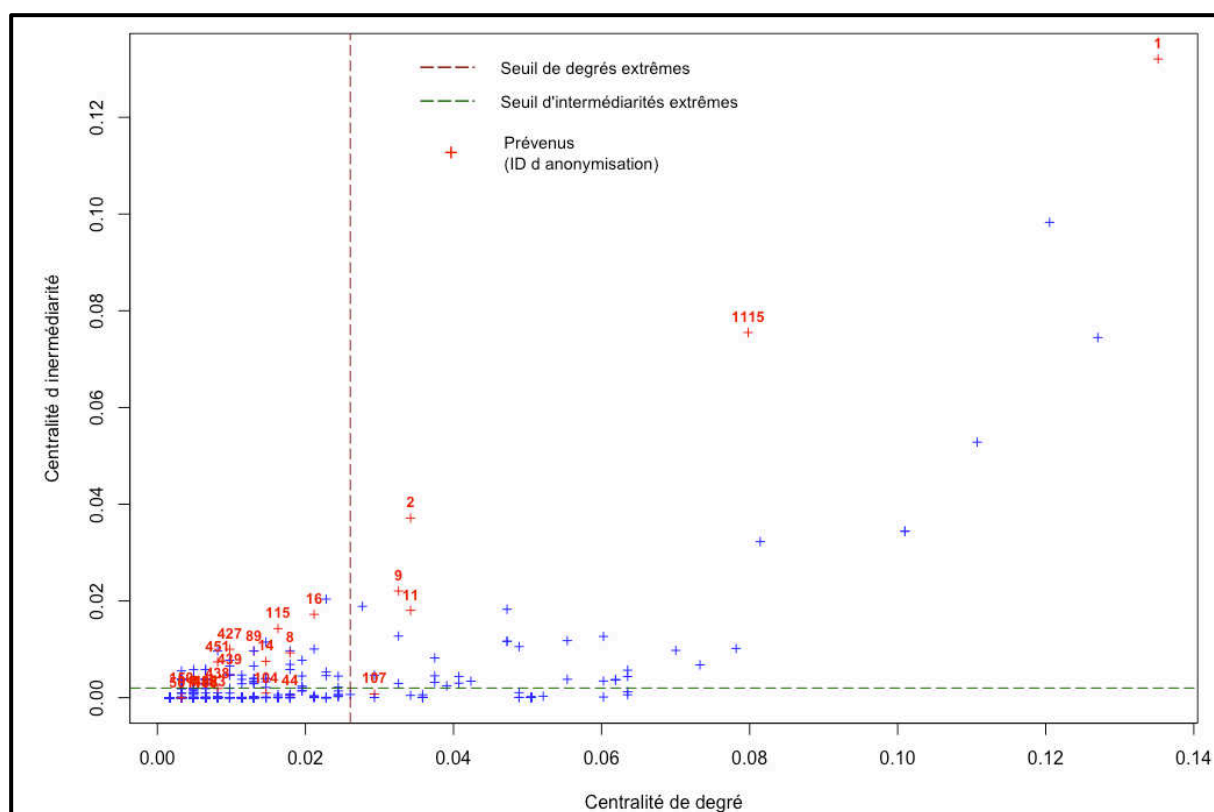


Figure 98 Scatterplot représentant les positions des acteurs dans l'espace graphique des centralités de degré et d'intermédiarité.

⁴⁷³ Cf. section VII.1 a) « Traitements préliminaires à l'analyse ».

⁴⁷⁴ Le terme « primaire » est toujours employé par opposition au terme « secondaire » par souci de clarté. Lorsqu'aucun de ces termes n'est mentionné, nous parlons par défaut de centralités « primaires ».

⁴⁷⁵ *Idem*

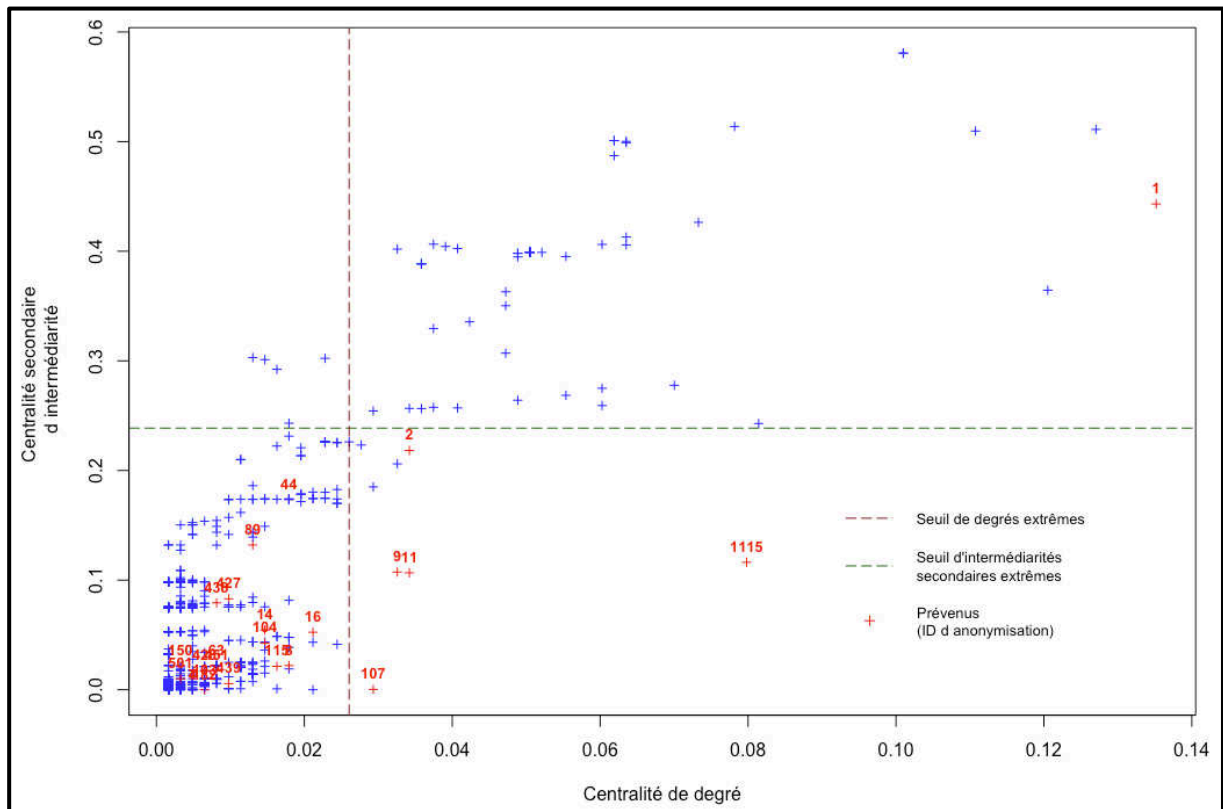


Figure 99 Scatterplot représentant les acteurs en fonction de leurs centralités primaire de degré et secondaire d'intermédiarité. Sont mis en évidence en rouge les prévenus désignés par leur identifiant d'anonymisation.

Tableau 68 Liste des prévenus (identifiants d'anonymisation) pour chacun des quatre profils de centralité étudié

Profil 3	Profil 4	Profil 7	Profil 8
Fort degré	Faible degré	Fort degré	Faible degré
Forte intermédiarité	Forte intermédiarité	Forte intermédiarité 2nd	Forte intermédiarité 2nd
1	16	1	
2	14		
9	89		
11	104		
1115	115		
107	150		
	8		
	427		
	451		
	438		
	439		

Sur les 25 prévenus mentionnés dans les données, 5 sont retrouvés dans le profil 3, 11 dans le profil 4, 1 dans les profils 3 et 7, et aucun dans le profil 8⁴⁷⁶.

⁴⁷⁶ Ce qui n'est pas forcément étonnant, dans la mesure où, comme explicité plus haut, concernant les profils 4 et 8, l'attendu n'est pas tant de valider les résultats analytiques en démontrant que l'enquête a effectivement révélé l'importance de ces individus, que de proposer une nouvelle approche pouvant guider les investigations futures sur les réseaux criminels.

Le tableau 69 résume les résultats des comparaisons entre les informations mentionnées dans les pièces récapitulatives du dossier d'enquête et les hypothèses de positionnement des acteurs dans les espaces de centralité émises à la section VII.4.

Compte tenu du volume de données que représente le dossier dans son intégralité (plus de 20 kg de documents, recto-verso), les hypothèses de positions n'ont été émises que pour les acteurs mentionnés dans les pièces récapitulatives.

Tableau 69 Résumé des comparaisons entre les résultats analytiques et les conclusions de l'enquête.

Accusé	Mention dans le résumé de l'enquête	Informations mentionnées dans les pièces récapitulatives	Hypothèse de position émise	Position	Hypothèse confirmée
1	OUI	Collecteur principal sous les ordres des <i>sarafs</i> belges	Profil(s) 3 et/ou 7	Profils 3 et 7	OUI
2	OUI	Collecteur sous les ordres des <i>sarafs</i> belges	Profil(s) 3 et/ou 7	Profil 3	OUI
9	OUI	<i>Saraf</i> parisien (patron)	Profil(s) 3 et/ou 7	Profil 3	OUI
11	OUI	<i>Saraf</i> parisien (employé)	Profil(s) 3 et/ou 7	Profil 3	OUI
107	OUI	Collecteur sous les ordres des <i>sarafs</i> marocains	Profil(s) 3 et/ou 7	Profil 3	OUI
1115	NON ⁴⁷⁷	Exerçait une activité illicite pour le compte de son père, Christian Ruffaut ⁴⁷⁸ demeurant au Maroc	NON	Profil 3	
16	NON		NON	Profil 4	
14	NON		NON	Profil 4	
89	NON		NON	Profil 4	
104	NON		NON	Profil 4	
115	NON		NON	Profil 4	
150	NON		NON	Profil 4	
8	NON		NON	Profil 4	
427	NON		NON	Profil 4	
451	NON		NON	Profil 4	
438	NON		NON	Profil 4	
439	NON		NON	Profil 4	

Les quatre hypothèses que nous avons émises, sans connaissance des identifiants des acteurs évoqués dans les pièces récapitulatives, et donc sans connaissance des positions de ces

⁴⁷⁷ Un dossier entier, évoqué dans le chapitre III, lui a été consacré dans le cadre de l'investigation. Les identités données dans le manuscrit sont fictives.

⁴⁷⁸ Pseudonyme attribué dans le respect de l'anonymisation en vertu de la circulaire ministérielle n°30053 du 3 septembre 2019.

derniers dans les espaces de centralités, et que nous avons pu confronter aux données sont validées. Ce résultat, bien que ténu, est encourageant.

CHAPITRE VIII INTERPRÉTATIONS EN TERMES D'ESPACE ET DE CHAMP CRIMINELS : LE COUP DE TONNERRE DU 10 NOVEMBRE

Après comparaison avec le réseau INR (extrait depuis la matrice intitulée « base complète » fournie par la SR de Paris XX^{ème}), nous avons fait le 10 novembre 2022, à J-15 du dépôt du manuscrit, les 8 constats suivants⁴⁷⁹ :

- Si certains membres de la famille M. sont renseignés dans les données, aucune mention n'est faite des liens de parenté qui les unissent ;
- Aucun des membres de la famille B. n'est renseigné dans la matrice de données. La seule mention que nous avons en pu trouver est contenue dans la colonne « Divers » de l'un des individus recensés dans les données : « *En lien téléphonique avec T.B.* ». Aucune entité ne représente ni la famille B., ni aucun de ses membres ;
- Concernant les clients désignés par des codes, les seules mentions dans les données sont une entité « code 25 » (2 occurrences), et les deux libellés de liens suivants : « *contact téléphonique au sujet femme code 25* » et « *contact pour un code 42* » ;
- Seul deux des trois collecteurs sous les ordres des *sarafs* belges existent dans les données ;
- Aucune mention n'est faite du lieu de stockage des fonds en Belgique ;
- L'adresse de l'épicerie du 17ème n'est pas renseignée dans les données. Cette épicerie, pourtant centrale dans les activités, n'a que 15 occurrences dans les données (le réquisitoire définitif mentionnant « *de très nombreuses allées et venues* ») ;
- Aucune mention des Pays-Bas, des EAU, de l'investisseur ou du trader identifiés ni des autres traders œuvrant pour A.M.

En conclusion, la plupart des informations concernant le « réseau criminel » décrit sont absentes des données brutes initiales. Il s'agit là pour nous d'un « coup de tonnerre » qui intervient en toute fin de rédaction de thèse. En outre, sur les 36 personnes contre qui une information a été suivie, 11 n'existent pas dans les données (ni entité représentative ni mention en attributs d'entités ou dans les libellés de liens), soit 30,6% des individus.

⁴⁷⁹ Cf. section III.2 a) « Résumé de l'enquête ».

Après 5 années d'études, d'abord en physique appliquée puis en sciences criminelles, à apprendre l'importance de la rigueur et de la qualité de la saisie et la structuration des données, nous n'avons pas pris conscience, lors du travail sur des données cette fois réelles et opérationnelles, des réalités du terrain, de l'ampleur des contraintes temporelles et opérationnelles de l'analyste, qui diffèrent de celles rencontrées dans le cadre de la formation académique. Par ailleurs, l'investigation ayant été réalisée par les SR de Paris XX^{ème} et de Marseille, peut-être existe-t-il un deuxième fichier relationnel, distinct de celui fourni par la SR de Paris, contenant les données relationnelles manquantes pourtant collectées lors de l'investigation. De par la rigueur inculquée par le parcours académique effectué avant la thèse, nous n'avons envisagé cette possibilité à aucun moment.

Il ne s'agit pas ici de dire que les données ne sont pas exploitables (pour atteindre les objectifs de recherche que nous nous étions fixés) car les personnes stratégiques n'y sont pas mentionnées – nous avons d'ailleurs avancé au chapitre VII que les personnes les plus stratégiques pour le réseau sont protégées par la structure de l'organisation, de telle sorte qu'elles soient peu visibles, et ainsi difficilement détectables. Nous soulignons le fait que les personnes stratégiques sont d'autant moins détectables si les personnes identifiées comme ayant des fonctions essentielles – et donc les liens qu'elles ont avec les autres acteurs de l'organisation – sont absentes des données à disposition.

La déception de se rendre compte non pas de l'incomplétude des données (par définition, la donnée est incomplète, l'enquêteur n'étant pas omniscient), mais du manque crucial des informations essentielles pourtant connues à J - 15 du rendu du manuscrit est au-delà des mots. Après 41 mois de travail, je réalise que les données sont inexploitable et les méthodes développées sont inapplicables car les résultats en seraient non contrôlés (dans le sens où l'on sait qu'ils ne seraient absolument pas cohérents ni représentatifs de l'enquête). Malgré les efforts, la volonté de résoudre les problèmes administratifs et techniques qui se sont posés ainsi que la volonté de faire avancer l'état des connaissances en proposant une méthode intéressante et en ouvrant de nouvelles pistes de recherche, c'est avec un sentiment d'échec traumatisant que s'achève cette thèse. Cette découverte est vécue comme un retour à la case départ, sans possibilité de reprendre depuis le début, de recommencer ou de tester d'autres idées. Seule reste la satisfaction d'avoir, tout au long de la thèse, suivi un processus d'analyse scientifique et rigoureux, sans jamais consulter les conclusions de l'enquête, en ne travaillant qu'à l'aveugle sur la base des données brutes telles que fournies par la SR de Paris afin d'éviter

tout biais d'analyse, tel que l'effet tunnel⁴⁸⁰. Arriver à la connaissance, quitte à ce que ce soit la connaissance d'une méconnaissance, est l'objet de cette thèse, bien que cela soit peu « consolant ». Il est temps pour moi de reprendre le « je » dans la thèse : je ne peux pas cacher que ce coup de tonnerre est terrible pour moi. Il reste que je n'ai pas triché et que travailler à l'aveugle était et demeure pour moi un aspect majeur de la thèse, garant de la solidité du raisonnement scientifique. Le prix est lourd.

Toutes les idées de travailler sur les notions d'espace social criminel, de champ criminel et de capitaux sociaux sont devenues irréalisables dès le moment où j'ai compris que la validité des résultats analytiques ne serait ni démontrable, ni même quantifiable... J'ai beaucoup de mal à concevoir la recevabilité scientifique de résultats issus d'un travail en mode dégradé, tant la qualité de la donnée s'avère défailante par rapport aux objectifs de la recherche (en termes de volume, de précision, de rigueur et de standardisation de la saisie, incomplétude des informations pourtant à disposition, d'absence d'entités, de liens et d'attributs fondamentaux pour une analyse pertinente). A mon sens, la rigueur scientifique m'interdit, avec de telles incertitudes, de tirer des conclusions des résultats exposés.

Le lecteur pourrait avancer qu'il existe en informatique des techniques dites « de reconstitution de données manquantes »⁴⁸¹. Ces techniques nécessitent tout de même de disposer d'une base depuis laquelle reconstituer les informations manquantes. Dans notre cas, lorsque ni les entités ni les liens ne sont renseignés, elles sont très difficilement applicables, même si nous disposions encore de quelques mois pour reprendre la recherche, mois que nous n'avons de toute façon malheureusement pas.

J'émetts cependant avant de conclure cette thèse une dernière idée, que je propose d'explicitier en reprenant et développant l'analogie explicative que j'avais utilisée lors de ma participation aux épreuves intra école doctorale⁴⁸² puis intra régionales⁴⁸³ au concours « Ma thèse en 180 secondes » au printemps 2022⁴⁸⁴.

Les centralités d'un acteur, au-delà de l'importance structurelle pour le réseau, peuvent également être pensées en termes d'avantages individuels résultant de cette position du point

⁴⁸⁰ Patrick Perrot, « L'analyse du risque criminel : l'émergence d'une nouvelle approche », *Revue de l'électricité et de l'électronique*, vol. 8, , 2014, pp. 100-107.

⁴⁸¹ Annamaria Ficara, Lucia Cavallaro, Francesco Curreri, Giacomo Fiumara, Pasquale De Meo, Ovidiu Bagdasar, Wei Song et Antonio Liotta, « Criminal Networks Analysis in Missing Data scenarios through Graph Distances », *PLOS ONE*, vol. 16, n° 8, 2021, pp. 1-18.

⁴⁸² École Doctorale Montaigne Humanité, Université Bordeaux Montaigne.

⁴⁸³ Région Aquitaine, Universités de Pau, La Rochelle, Bordeaux et Bordeaux Montaigne.

⁴⁸⁴ Présentation disponible à l'adresse : <https://www.youtube.com/watch?v=BhhRItr8Bjc&t=111s>.

de vue de l'acteur lui-même. Typiquement, un grand nombre de connecteurs dans un réseau est signe de compartimentalisation⁴⁸⁵, de sophistication, de flexibilité et de résilience⁴⁸⁶, avantage bénéficiant à l'organisation criminelle, tandis qu'une position de connecteur confère importance, contrôle des flux et ainsi une position stratégique aux acteurs qui les détiennent.

Dans la section II.3 b), une distinction entre un réseau en tant que structure ou en tant qu'acteur est exposée. D'une part, le réseau comme acteur est une forme d'action coordonnée et collective. C'est – comme son nom l'indique – un acteur en soi : les acteurs individuels sont liés par la possibilité d'agir collectivement, des objectifs communs, une action volontaire et stratégique. D'autre part, le réseau comme structure est un réseau qui n'existe que par sa structure formelle : un réseau social relie des protagonistes qui partagent un milieu pratique ou un certain type d'activité, mais dont les choix ne sont pas coordonnés⁴⁸⁷. Les individus membres de ces réseaux ne sont pas caractérisés par des intérêts, des stratégies ou des objectifs communs, mais par le fait d'être liés *de facto* par les mêmes activités sociales au sein d'un même champ.

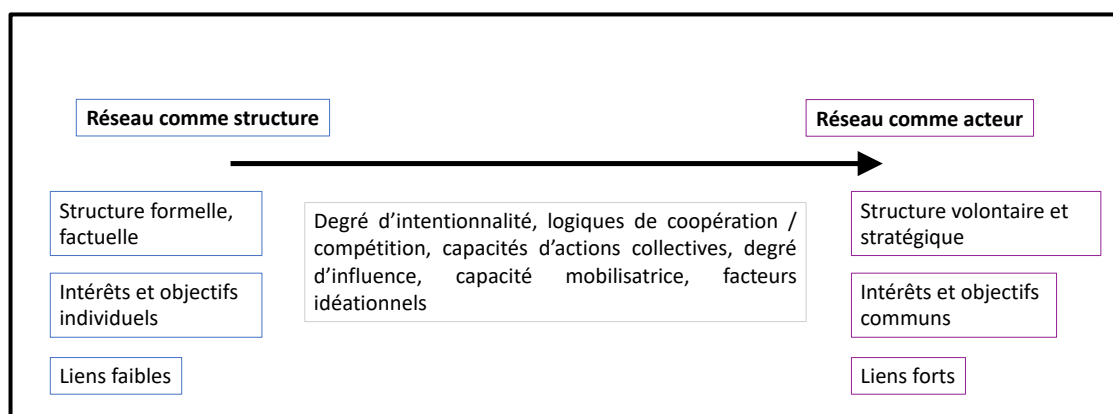


Figure 100 Continuum du réseau comme structure au réseau comme acteur

Être à même de comprendre si les bénéfices des positions stratégiques reviennent aux acteurs à titre individuel ou au réseau dans son ensemble pourrait permettre de positionner le réseau sur le continuum « structure-acteur » (figure 100), et donc de mieux interpréter sa structure. A l'inverse, positionner le réseau sur le continuum aidera à analyser les positions des acteurs pour déterminer leur niveau d'implication dans les activités criminelles étudiées.

⁴⁸⁵ United States DoD, *Countering Threat Networks*, *op. cit.*

⁴⁸⁶ Peter Klerks, « The Network Paradigm Applied to Criminal Organisations », *op. cit.* ; Geoffrey Pearson et Dick Hobbs, *Middle Market Drug Distribution*, *op. cit.* ; Nigel Coles, « It's not what you know— it's who you know that counts: analysing serious crime groups as social networks », *The British Journal of Criminology*, vol. 41, n° 4, 2001, pp. 580-594.

⁴⁸⁷ Amélie Forget, « Du réseau comme structure au réseau comme acteur : comment les réseaux sociaux se constituent en acteurs collectifs », *op. cit.*

C'est pour expliciter cette distinction qu'intervient l'analogie du jeu d'échecs. Le texte suivant a été écrit pour le concours « Ma thèse en 180 secondes »⁴⁸⁸ et développé comme fil rouge cognitif dans cette thèse. Dans le cas d'un réseau-acteur, chaque couleur revient à un joueur, dont l'objectif est de remporter la partie en utilisant l'ensemble de ses pièces et en maximisant leur efficacité collective, quitte à sacrifier plusieurs pièces. Dans le cas d'un réseau-structure, chaque pièce est un joueur à part entière, dont l'objectif est le maintien d'une position avantageuse, la survie individuelle tout au long de la partie.

Analogie du jeu d'échecs : ma thèse en 180 secondes et apports réflexifs

« Lorsque vous jouez aux échecs, vous préférez être les blancs ou les noirs ? J'ai toujours préféré être les noirs, pour adapter ma stratégie à celle de mon adversaire, contrer ses mouvements. Mon adversaire ? Le réseau criminel que j'affronte. Ma stratégie ? Détruire la sienne.

L'échiquier est vaste : c'est un plateau social à bien plus de deux dimensions sur lequel évoluent un grand nombre d'acteurs sociaux. Si les pions des forces de l'ordre sont relativement connus de l'adversaire (les règles selon lesquelles jouent les forces de l'ordre sont délimitées et encadrées par un cadre légal et juridique), la réciproque est fautive : à chaque nouvelle partie se créent un nouvel échiquier, de nouvelles pièces, de nouvelles règles, et de nouvelles stratégies – par définition, ni les organisations criminelles ni leurs pions ne s'en tiennent aux règles du jeu.

Certes, les règles du jeu sont censées être établies : les lois. Mais par définition, les criminels refusent de s'y plier – ils imposent leurs règles. Or, en matière de criminalité et pour mon adversaire, gagner signifie simplement ne pas perdre. Tant que le réseau n'est pas en échecs et mat, il reste actif, et la partie continue.

Comment appréhender efficacement la stratégie adverse à chaque nouvelle partie ?

1 – En comprenant les règles selon mon adversaire joue ;

2 – En identifiant ses pièces maîtresses, et en les sortant de l'échiquier

Comprendre les règles du jeu, c'est déterminer la structure sociale, externe et interne, du réseau criminel : est-il dense ou éparé ? Centralisé ou hiérarchisé ? Connecté ou éclaté ? Acteur uni ou structure non coordonnée ?

Identifier les pièces maîtresses, c'est comprendre qui détient le pouvoir, l'influence, la capacité d'action, qui contrôle les flux d'informations.

⁴⁸⁸ <https://mt180.fr/>

Et quels sont les capitaux sociaux qui leur permettent de se maintenir à ces positions ? La Reine est puissante car elle se déplace comme elle le veut, les cavaliers sont imprévisibles car ils sautent des cases. Dans certaines configurations, les fous, qui se déplacent en diagonale, sont plus stratégiques que les tours qui sont mobiles en ligne droite ; dans d'autres situations ce sera l'inverse.

Pour déterminer cela, j'analyse les positions relatives des membres du réseau, et je mesure, selon plusieurs critères, à quel point chaque pièce est importante pour le bon fonctionnement de la stratégie de mon adversaire. Je peux alors identifier ses fous, ses cavaliers, sa reine, et leurs possibilités de se déplacer sur l'échiquier – les atouts qu'il détiennent. Je sais maintenant quelles pièces, quels atouts retirer du plateau pour pénaliser mon adversaire et entraver sa stratégie. Bien sûr, les forces de l'ordre n'en sont pas à leur première partie : de la même manière que les joueurs d'échecs professionnels maîtrisent parfaitement les grandes ouvertures et leurs contre-mouvements, les policiers et gendarmes savent appliquer les contre-mesures efficaces contre les stratégies criminelles les plus courantes... Mais qui évoluent rapidement. Cette thèse n'a pas pour prétention de leur apprendre à jouer, mais cherche à développer de nouvelles méthodes pour appréhender et reconnaître les stratégies criminelles émergentes, anticiper les coups à jouer pour empêcher l'adversaire de mettre en place sa stratégie, plutôt que d'y réagir au coup par coup. Les forces de l'ordre auraient ainsi plusieurs coups d'avance pour pénaliser et neutraliser les pièces maîtresses des réseaux criminels, jusqu'à les mettre en échec et mat. »

Dans cette perspective, les positions des pièces sur l'échiquier doivent-elles être analysées du point de vue individuel : individus servant leurs propres intérêts et objectifs, s'associant dans le but de maximiser leurs profits personnels, et néanmoins en concurrence pour les meilleures places de l'échiquier ? Ou plutôt dans leur ensemble, comme un réseau qui place ses pions pour mettre en place une stratégie globale, unifiée, en disposant ses pièces de façon tactique, en fonction de leurs atouts pour maximiser pour optimiser leur efficacité collective ?

Faire du renseignement criminel, cela peut être comme étudier les ouvertures les plus célèbres, les plus courantes, et leurs contre mouvements. C'est aussi déterminer selon quelles règles implicites joue l'adversaire (consciemment ou inconsciemment), si certaines règles sont particulièrement appliquées, récurrentes. Si on imaginait pouvoir étudier son adversaire aux échecs dans le but d'identifier avant même de commencer la partie les stratégies qu'il va mettre en œuvre (et les règles selon lesquelles il va jouer), on pourrait remporter la partie très tôt, en

anticipant les coups à jouer pour empêcher l'adversaire de mettre en place sa stratégie, plutôt que de réagir à ses mouvements au coup par coup.

CONCLUSION

Peut-on combiner les outils analytiques issus de diverses disciplines pour développer, sur la base de données telles que collectées et saisies lors des enquêtes de gendarmerie, une méthodologie de représentation, de modélisation et d'analyse de réseaux criminels aboutissant à une détection des fonctions et positions stratégiques dans le fonctionnement de l'organisation ?

Telle est la question à laquelle nous avons tenté de répondre au cours de la thèse. Y apporter des éléments de réponse aura été un voyage au travers de diverses disciplines, une découverte de paysages et horizons analytiques, tant en termes de rapport à la réalité du terrain et aux contraintes opérationnelles que d'approche, de méthodologie, d'outils, de connaissances et de solutions.

Nous avons, dans un premier chapitre, abordé en détail la genèse de ce projet de doctorat, au travers des différents partenariats qui lui ont donné vie et de son aspect multidisciplinaire. Dans un deuxième chapitre, l'état d'avancement des connaissances dans les différentes disciplines auxquelles fait appel ce travail a été exposé. La rédaction de cet état de l'art a guidé la manière dont ont été appréhendées les interactions entre les disciplines pour produire du renseignement criminel.

Le cas d'usage étudié, les données à disposition et leur structure ainsi que les prétraitements appliqués pour préparer et formater la donnée ont fait l'objet des troisième et quatrième chapitres. La structure de la donnée, collectée et saisie par et pour l'enquête judiciaire, a été un élément majeur dans l'évaluation de son exploitabilité par le renseignement criminel, notamment au travers de la méthode que nous avons développée.

Le cinquième chapitre s'est attaché à décrire l'ensemble des étapes de transformation de la donnée qui ont conduit à l'établissement du réseau social criminel que nous avons par la suite analysé. L'objet du sixième chapitre a été de discuter cette méthode, d'en envisager les biais et d'en qualifier les performances. Le septième chapitre a été consacré dans un premier temps à la description, en termes structurels, du réseau criminel obtenu, et dans un second temps à son analyse hybride, tant depuis les perspectives de l'analyse de réseaux sociaux et de la théorie des graphes que du renseignement criminel que nous cherchions à en extraire. Enfin, le huitième chapitre, peu conventionnel, s'est autorisé une narration de la fin de thèse, depuis la

perception de la doctorante – ma perception, et est revenu sur les difficultés liées à l’analyse de données interdépendantes qui constituent un ensemble relationnel.

L’intégralité de la démarche analytique a été effectuée à l’aveugle, c’est-à-dire que bien que nous disposions de ces informations depuis juin 2020, les identités des protagonistes ainsi que les résumés et conclusions de l’enquête nous sont restées inconnus jusqu’au moment de comparer nos résultats avec ceux de l’enquête.

Le premier résultat que nous avons proposé est un sociogramme représentant tous les acteurs graphiquement connectés sur la base des données brutes, c’est-à-dire pour lesquels les données initiales permettaient, en suivant une succession continue d’entités (de nœuds) et de liens (d’arêtes) de générer un lien direct. Pour chacun de ces liens, le niveau de confiance que nous pouvions accorder à son existence sur le terrain, *i.e.* au fait que les acteurs ainsi connectés étaient effectivement en relation dans le cadre de l’enquête a été quantifié à l’aide d’une pondération. Cette pondération chiffrée a été obtenue sur une base sémantique, en s’inspirant de concepts issus des grammaires formelles pour développer un modèle (semi) automatisé de mise en évidence de liens directs entre individus dissimulés, disséminés, dans un ensemble conséquent de données de diverses natures. Une automatisation complète n’a pas été envisagée dans la mesure où le raisonnement et l’expérience de l’analyste sont et restent une composante essentielle de l’analyse criminelle.

Nous avons par la suite utilisé des outils analytiques pour déterminer comment interpréter ces pondérations et fixer un seuil d’acceptation des liens directs sur la base de la capacité de ces pondérations à refléter le caractère hypothétique ou non des liens directs entre personne. Deux objectifs du renseignement criminel, à savoir le soutien à l’enquête judiciaire et la capitalisation de connaissances par les forces de l’ordre, ainsi que les contraintes opérationnelles qui y sont liées, ont été considérés ; deux seuils distincts adaptés aux différents besoins ont été définis. Le sociogramme a été filtré à l’aide du seuil proposé dans une approche de renseignement criminel pour la détection et la compréhension de mécanismes.

En effet, il existe de nombreuses manières, reflétées par différents indicateurs, d’estimer la performance d’un modèle, par définition imparfait, et de prioriser les aspects de la réalité que le modèle doit incarner. Dans une logique de renseignement criminel pour l’analyse de phénomènes, l’analyste est souvent amené à traiter plusieurs affaires afin de détecter des tendances dans les schémas sociaux, fonctionnels et organisationnels récurrents, permettant

ainsi de capitaliser des connaissances sur les mécanismes sociaux à l'œuvre dans un champ criminel donné. Dans de tels cas, le volume de données à disposition rend inenvisageable les contrôles manuels. Nous proposons ainsi un seuil minimisant le taux d'erreur global et maximisant la précision du modèle, c'est-à-dire que nous acceptons de rejeter des liens que l'analyste aurait pourtant gardé s'il avait été possible de tous les analyser manuellement, en ayant cependant un niveau de confiance solide dans l'existence des liens conservés par le modèle. Les tests de performance effectués indiquent que ce seuil confère au modèle une erreur globale de 8% et une précision de 90%.

Le réseau social de l'investigation ainsi obtenu, modélisé par le graphe *Investigation_Network_Rens (INR)*, contient 615 individus, reliés par un total de 2 065 relations. Ce réseau criminel a été analysé en étudiant d'abord sa structure, sa connectivité, son niveau de cohésion, puis le positionnement des acteurs les uns par rapport aux autres, notamment au travers de l'analyse de leurs centralités de degré et d'intermédiarité. Les mesures usuelles de ces deux formes de centralité ont été appliquées, et deux nouveaux indicateurs, quantifiant ce que nous avons appelé les centralités secondaires ont été proposés. Ces derniers sont inspirés d'une mesure existante appelée centralité de vecteur propre, qui permet de mesurer le niveau de connexion d'un acteur aux autres acteurs importants du réseau. Nous avons alors redéfini le concept d'importance d'une part selon la centralité de degré, mesurant le nombre de relations directes d'un individu, et d'autre part selon la centralité d'intermédiarité, quantifiant dans quelle mesure un acteur permet de relier, *via* des trajets courts, deux nœuds ou deux parties distinctes du réseau, parfois en étant un intermédiaire obligatoire. Les mesures de centralités secondaires ont ainsi permis de mesurer le niveau de connexion des acteurs du réseau aux acteurs identifiés comme importants selon les mesures de centralité primaires.

Les résultats ont montré que la centralité secondaire de degré ne permettait pas – hormis éventuellement dans de rares cas – d'identifier des acteurs que la centralité primaire de degré n'avait déjà mis en évidence. La centralité secondaire d'intermédiarité a montré un potentiel de détection complémentaire d'acteurs, par le fait qu'elle n'est pas significativement corrélée à la mesure de l'intermédiarité primaire. Cependant, malgré ce fait encourageant, elle n'a en l'occurrence pas apporté d'éléments nouveaux à l'analyse.

Nous avons proposé d'identifier les acteurs stratégiques en travaillant sur la combinaison de ces indicateurs, en distinguant des profils de centralité. Nous nous sommes alors intéressés aux acteurs ayant d'une part une forte centralité de degré et d'intermédiarité, et d'autre part aux acteurs moins visibles, mais néanmoins stratégiques : les acteurs ayant une

faible centralité de degré mais une forte centralité d'intermédiation. Après avoir pris connaissance du résumé de l'enquête, et ainsi des différents rôles mis en évidence par les investigations, nous avons émis un certain nombre d'hypothèses sur le positionnement des acteurs détenant les rôles mentionnés dans l'espace des centralités.

Enfin, nous avons pris connaissance des identités des protagonistes et avons comparé leurs positions effectives avec les positions envisagées dans nos hypothèses. C'est à ce moment qu'a été fait le constat qu'une majorité des informations décrites dans les pièces récapitulatives du dossier judiciaire n'existait pas dans les données sur lesquelles nous avons choisi de travailler, à savoir la matrice d'analyse saisie par l'analyste de la SR de Paris en charge du dossier. Nous n'allons pas revenir sur les conséquences de cette découverte, discutées dans le huitième et dernier chapitre, mais plutôt souligner le fait que malgré cela, lorsque les données le permettaient, les hypothèses émises ont été confirmées.

Bien que les résultats finaux soient ténus, ces recherches ouvrent de nombreuses pistes et montrent que la pluridisciplinarité a bien sa place pour répondre aux problématiques du renseignement criminel.

En termes de recherche, le modèle d'extraction du réseau social depuis le graphe d'enquête multivarié constitue une première pierre à l'édifice : bien que de nombreuses optimisations s'imposent, les résultats sont encourageants. Ces optimisations concernent l'établissement de critères supplémentaires, sur lesquels les analystes raisonnent mais actuellement difficilement formalisables et formatables. Ces critères seront alors à prendre en compte lors de la catégorisation des liens et entités ainsi que lors de l'attribution des pondérations. L'implémentation de nouveaux éléments dans l'algorithme d'extraction, tels que la considération des dates et heures des liens, permettra également d'augmenter la performance et la précision du modèle.

Cette approche prometteuse s'accompagne d'un changement de paradigme dans la saisie de la donnée initiale. Le modèle proposé a pour objectif de fonctionner sur des données telles que collectées par les enquêteurs de la Gendarmerie nationale, notre place n'est pas de discuter de la collecte des informations, mais de la manière dont ces dernières sont saisies dans une matrice d'analyse. Pour qu'un algorithme puisse appuyer l'analyse des données, que ce soit avec l'ambition d'orienter les enquêtes en cours ou futures, cette dernière doit être structurée, standardisée et homogénéisée, sans quoi l'automatisation de certains traitements ne peut être envisagée.

L'approche pluridisciplinaire a permis le développement de la méthode proposée en apportant des perspectives, des outils, des arguments, des méthodologies et des idées qui s'axent en compléments les uns des autres et se nourrissent mutuellement. Nous sommes convaincue de son bien-fondé et de son potentiel d'extraction et de gestion de connaissances dans le domaine du renseignement criminel. Un approfondissement de la recherche, notamment en complétant les données initiales pourra permettre, à notre sens, de discuter des concepts d'espaces, de champs et de capitaux criminels que nous avons évoqués au cours de la thèse.

BIBLIOGRAPHIE

Accardo Alain, *Introduction à une sociologie critique : lire Pierre Bourdieu*, Marseille, Agone, 2006, 382 p.

Adler Patricia et Adler Pierre, « Shifts and oscillations in deviant careers: The case of upper-level drug dealers and smugglers. », *Social Problems*, vol. 31, n° 2, 1983, pp. 195-207.

Aeppli Pierre, Ribaux Olivier et Summerfield Everett, *Decision making in policing: operations and management*, Lausanne, EPFL-Press, 2011, 228 p.

Åkerström Malin, *Crooks and squares: lifestyles of thieves and addicts in comparison to conventional people*, New Brunswick, N.J., Transaction Books, 1995, 265 p.

Angelier Clarisse et Courouble Jeanne, *Ces créateurs d'entreprises innovantes : quand doctorat se conjugue avec entrepreneuriat*, Paris, Eyrolles-Éd. d'Organisation, 2011, 156 p.

Antwi Bosiakoh Thomas, « Sutherland, Edwin H. (1883–1950) », dans *Encyclopedia of the Sciences of Learning*, Boston, MA, Springer US, 2012, pp. 3253-3255.

Association of Chief Police Officers, *Guidance on the National Intelligence Model*, National Centre for Policing Excellence, 2005, 213 p.

Auber David, Archambault Daniel, Bourqui Romain, Delest Maylis, Dubois Jonathan, Lambert Antoine, Mary Patrick, Mathiaut Morgan, Melançon Guy, Pinaud Bruno, Renoust Benjamin et Vallet Jason, « TULIP 5 », dans Reda Alhajj et Jon Rokne (dir.), *Encyclopedia of Social Network Analysis and Mining*, New York, NY, Springer New York, 2017, pp. 1-28.

Auber David, Chiricota Yves, Delest Maylis, Melançon Guy, Domenger Jean-Philippe et Mary Patrick, « Visualisation de graphes avec Tulip : exploration interactive de grandes masses de données en appui à la fouille de données et à l'extraction de connaissances », Namur, Belgique, 2007.

Audit Commission for Local Authorities and the National Health Service in England and Wales, *Helping with enquiries: tackling crime effectively*, London, HMSO, 1993, 65 p.

Baker Wayne E. et Faulkner Robert R., « The Social Organization of Conspiracy: Illegal Networks in the Heavy Electrical Equipment Industry », *American Sociological Review*, vol. 58, n° 6, 1993, pp. 837-860.

- Barlatier Jérôme, « Criminal Investigation and Criminal Intelligence: Example of Adaptation in the Prevention and Repression of Cybercrime », *Risks*, vol. 8, n° 3, 18/09/2020, pp. 99-109.
- Barlatier Jérôme, *Management de l'enquête et ingénierie judiciaire, recherche relative à l'évaluation des processus d'investigation criminelle*, Thèse de doctorat, École des Sciences Criminelles, Université de Lausanne, 2017, 475 p.
- Beare Margaret E., *Critical Reflections on Transnational Organized Crime, Money Laundering, and Corruption*, Toronto, University of Toronto Press, 2017.
- Beauguitte Laurent, « E comme ERGM », <https://groupefmr.hypotheses.org/2718>, consulté le 2 mars 2021.
- Beauguitte Laurent, « Graphes, Réseaux, Réseaux sociaux : vocabulaire et notation – Version 1 », 2010, <https://hal.archives-ouvertes.fr/halshs-00541898/>, consulté le 12 novembre 2020.
- Becker Gary, « Crime and Punishment: An Economic Approach », *Journal of Political Economy*, vol. 76, n° 2, 1968, pp. 169-217.
- Becker Howard S., *Faire preuve : des faits aux théories*, Paris, la Découverte, 2020, 288 p.
- Becker Howard S., *Outsiders: Studies in the Sociology of Deviance*, Glencoe, 1963, 224 p.
- Béland Daniel, « Insecurity, Citizenship, and Globalization: The Multiple Faces of State Protection », *Sociological Theory*, vol. 23, n° 1, 2005, pp. 25-41.
- Berlière Jean-Marc et Lévy René, *Histoire des polices en France : de l'Ancien régime à nos jours*, Paris, Nouveau monde éd, 2013.
- Bernier Isabelle, *Histoire : la police à Paris sous l'Ancien Régime*, <https://www.futura-sciences.com/sciences/questions-reponses/epoque-moderne-histoire-police-paris-sous-ancien-regime-12299/>, consulté le 30 mars 2022.
- Bertrand Daisy, Fluss Joel, Billard Catherine et Ziegler Johannes C., « Efficacité, sensibilité, spécificité : comparaison de différents tests de lecture », *L'Année psychologique*, vol. 110, n° 2, 2010, pp. 299-320.
- Bidart Claire, « Brève initiation à l'analyse des réseaux à partir de la sociologie » par Claire Bidart | IMéRA - Institut d'études avancées, <https://imera.univ-amu.fr/fr/media/breve-initiation-a-lanalyse-reseaux-a-partir-sociologie-claire-bidart>, consulté le 27 octobre 2019.

Blake Robert R., « Review of Who Shall Survive? », *Sociometry*, vol. 17, n° 1, 1954, pp. 77-91.

Blum Alain et Gribaudo Maurizio, « Des catégories aux liens individuels : l'analyse statistique de l'espace social », *Annales. Histoire, sciences sociales*, vol. 45, n° 6, 1990, pp. 1365-1402.

Blumstein Alfred, *Criminal careers and « career criminals »*, Washington, DC, National Acad. Press, 1986, 458 p.

Boivin Rémi et Morselli Carlo, *Les réseaux criminels*, Montréal, Les Presses de l'Université de Montréal, 2016, 319 p.

Bonache Adrien Bernard, « Puissance statistique des tests et date de publication : une corrélation trompeuse ? », *Comptabilité Contrôle Audit*, Tome 24, n° 1, 12/03/2018, pp. 13-41.

Bonacich Phillip, « Some unique properties of eigenvector centrality », *Social Networks*, vol. 29, n° 4, 10/2007, pp. 555-564.

Bonnewitz Patrice, *Premières leçons sur La sociologie de Pierre Bourdieu*, Paris, Presses Universitaires de France, 2005, 124 p.

Bonnewitz Patrice, *Pierre Bourdieu : vie, œuvres, concepts*, Paris, Ellipses, 2002, 94 p.

Borgatti Stephen P., Everett Martin G. et Johnson Jeffrey C., *Analyzing social networks*, Los Angeles, SAGE, 2018, 363 p.

Bourdieu Pierre, *Raisons pratiques : sur la théorie de l'action*, Paris, Seuil, 1996, 251 p.

Bourdieu Pierre, *Choses dites*, Paris, Éditions de Minuit, 1987, 229 p.

Bourdieu Pierre, « The forms of Capitals », dans *Handbook of theory and research for the sociology of education*, Greenwood Press, 1986, pp. 241-258.

Bourdieu Pierre, « Espace social et genèse des "classes" », *Actes de la recherche en sciences sociales*, vol. 52-53, 1984, pp. 3-14.

Bourdieu Pierre, « Le capital social : Notes provisoires », *Actes de la recherche en sciences sociales*, vol. 31, n° 1, 1980, pp. 2-3.

Bourdieu Pierre, « Les trois états du capital culturel », *Actes de la Recherche en Sciences Sociales*, vol. 30, 1979, pp. 3-6.

Bourgois Philippe, *In search of respect: selling crack in El Barrio*, Cambridge, Cambridge University Press, 2014, 434 p.

Bouroche Jean-Marie et Saporta Gilbert, *L'analyse des données*, Paris, Presses Universitaires de France, 2005, 128 p.

Brandes Ulrik, « A faster algorithm for betweenness centrality* », *The Journal of Mathematical Sociology*, vol. 25, n° 2, 06/2001, pp. 163-177.

Bright David A., Greenhill Catherine, Reynolds Michael, Ritter Alison et Morselli Carlo, « The Use of Actor-Level Attributes and Centrality Measures to Identify Key Actors: A Case Study of an Australian Drug Trafficking Network », *Journal of Contemporary Criminal Justice*, vol. 31, n° 3, 2015, pp. 262-278.

Bright David A., Brewer Russell et Morselli Carlo, « Using social network analysis to study crime: Navigating the challenges of criminal justice records », *Social Networks*, vol. 66, , 01/07/2021, pp. 50-64.

Bunnin F. O., Shenvi Aditi et Smith Jim Q., « Network Modelling of Criminal Collaborations with Dynamic Bayesian Steady Evolutions », *arXiv:2007.04410 [cs, stat]*, 08/07/2020.

Burt Ronald S., *Brokerage and closure: an introduction to social capital*, Oxford, Oxford University Press, 2005, 279 p.

Burt Ronald S., « Le capital social, les trous structureaux et l'entrepreneur », *Revue Française de Sociologie*, vol. 36, n° 4, 10/1995, pp. 599.

Burt Ronald S., *Structural holes: the social structure of competition*, Cambridge, Mass., Harvard Univ. Press, 1995, 313 p.

Campana Paolo, « Explaining criminal networks: Strategies and potential pitfalls », *Methodological Innovations*, vol. 9, 2016, pp. 1-10.

Caplain Serge, « Penser son ennemi. Modélisations de l'adversaire dans les forces armées », *IFRI*, n° 82, 2018, pp. 66.

Cavallaro Lucia, Bagdasar Ovidiu, De Meo Pasquale, Fiumara Giacomo et Liotta Antonio, « Graph and Network Theory for the analysis of Criminal Networks », *arXiv:2103.02504 [physics]*, 2021, pp. 139-156.

- Champin Christophe, *Afrique noire, poudre blanche : l'Afrique sous la coupe des cartels de la drogue*, Bruxelles, A. Versailles, 2010, 155 p.
- Charette Yanick, *La perception du prestige des occupations illicites par des délinquants : une perspective sur les capitaux criminels*, Mémoire de maîtrise, Université de Montréal, 2010, 95 p.
- Charpier Frédéric, *Islamisme, 60 ans de lutte antiterroriste*, Paris, Tallandier, 2017, 333 p.
- Chassaigne Philippe, « Criminalité et police à Londres au XIXe siècle », *Revue française de civilisation britannique*, vol. 12, n° 3, 2003, pp. 2-7.
- Chikhi Nacim Fateh, *Calcul de centralité et identification de structures de communautés dans les graphes de documents*, Université Paul Sabatier - Toulouse III, Toulouse, 2010, 182 p.
- Clarke Ronald V et Eck John E, *Crime Analysis for Problem Solvers in 60 Small Steps*, Washington, DC, U.S. Department of Justice, COPS, 2005, 150 p.
- Cloward Richard A., « Illegitimate Means, Anomie, and Deviant Behavior », *American Sociological Review*, vol. 24, n° 2, 1959, pp. 164-176.
- Cloward Richard A. et Ohlin Lloyd, *Delinquency and opportunity*, Free Press, 1966, 230 p.
- Cohen Albert K. , *Delinquent Boys: The Culture of the Gang*, New York, The Free Press, 1955, 178 p.
- Cohen Jacob, *Statistical power analysis for the behavioral sciences*, Hillsdale, N.J, L. Erlbaum Associates, 1988, 567 p.
- Cole Simon A., *Suspect identities: a history of fingerprinting and criminal identification*, Cambridge, MA, Harvard Univ. Press, 2002, 369 p.
- Coles Nigel, « It's not what you know— it's who you know that counts: analysing serious crime groups as social networks », *The British Journal of Criminology*, vol. 41, n° 4, 2001, pp. 580-594.
- Colin G. G., Aitken et Taroni Franco, *Statistics and the evaluation of evidence for forensic scientists*, Chichester, England ; Hoboken, N.J, Wiley, 2004, 509 p.
- Colquhoun Patrick (1800), *A Treatise on the Commerce and the Police of the River Thames*, Montclair, Patterson Smith, 1969, 712 p.

Colquhoun Patrick (1806), *Treatise on the Police of the Metropolis*, Montclair, Patterson Smith, 1969, 519 p.

Cornish D. B., « Theories of action in criminology: Learning theory and rational choice approaches », *Routine activity and rational choice*, 1993, pp. 351-382.

Cornish D. B., « The procedural analysis of offending and its relevance for situational prevention », *Crime prevention studies*, vol. 2, n° 1, 1994, pp. 32-78.

Coyne Michelle A. et Eck John E., « Situational Choice and Crime Events », *Journal of Contemporary Criminal Justice*, vol. 31, n° 1, 2015, pp. 12-29.

Cucherat Michel, Lièvre Michel, Leizorovicz Alain et Boissel Jean-Pierre, *Lecture critique et interprétation des résultats des essais cliniques pour la pratique médicale*, Paris, Flammarion médecine-sciences, 2004.

Cullen Francis T. et Link Bruce G., « Crime as an occupation. A New Look », *Criminology*, vol. 18, n° 3, 1980, pp. 399-410.

Maurice Cusson, *Répétitions criminelles, renseignements et opérations coup-de-poing*, Chicoutimi, Bibliothèque québécoise, 2012, 23 p.

Cusson Maurice, *L'art de la sécurité : ce que l'histoire de la criminologie nous enseigne*, Lausanne, Presses polytechniques et universitaires Romandes, 2011, 248 p.

Cusson Maurice, *La délinquance, une vie choisie : entre plaisir et crime*, Chicoutimi, Bibliothèque québécoise, 2008, 34 p.

Cusson Maurice, *Délinquants, pourquoi ?*, Montréal, Bibliothèque québécoise, 1995, 238 p.

Cusson Maurice, Lemieux Frédéric et Dupont Benoît, *Traité de sécurité intérieure*, Lausanne, Presses polytechniques et universitaires romandes, 2007, 705 p.

De Larmina Xavier, *Sociologie de la déviance : des théories du passage à l'acte à la déviance comme processus*, <http://ses.ens-lyon.fr/articles/sociologie-de-la-deviance>, consulté le 6 mai 2020.

De Maillard Clément, « La France et le renseignement criminel : entre volonté et réalité, une ambition à écrire », *Sécurité et stratégie*, vol. 17, n° 2, 2014, pp. 49-59.

- Degenne Alain et Forsé Michel, *Les réseaux sociaux : une analyse structurale en sociologie*, Paris, A. Colin, 1994, 288 p.
- Delas Jean-Pierre et Milly Bruno, « Les individualismes méthodologiques », dans *Histoire des pensées sociologiques*, Paris, Armand Colin, 2015, pp. 371-411.
- Drevelle Matthieu, « C comme centralité(s) », <https://groupefmr.hypotheses.org/2324>, consulté le 5 novembre 2022.
- Durkheim Émile, *Les règles de la méthode sociologique*, Paris, Flammarion, 2010, 254 p.
- Durkheim Émile et Hocquard Anita, *Le suicide. Livre II*, Paris, GF Flammarion, 2014, 304 p.
- Durkheim Émile et Paugam Serge, *Le suicide : étude de sociologie*, Paris, PUF, 2013, 482 p.
- Durkheim Émile, Weber Florence, Berthelot Jean-Michel et Mucchielli Laurent, *Les règles de la méthode sociologique*, Paris, Flammarion, 2017, 150 p.
- Emsley Clive et Gervais Pierre, « Police, maintien de l'ordre et espaces urbains : une lecture anglaise », *Revue d'histoire moderne et contemporaine*, vol. 50, n° 1, 2003, pp. 5-12.
- Esseiva Pierre, Gaste Laeticia, Alvarez Daniel et Anglada Frederic, « Illicit drug profiling, reflection on statistical comparisons », *Forensic Science International*, vol. 207, n° 1, 2011, pp. 27-34.
- Faucon Benoît et Fayol Clément, *Un cartel nommé Daech : révélations sur les réseaux mafieux de l'État islamique*, Paris, First éditions, 2017, 230 p.
- Ferrand Alexis, « La structure des systèmes de relation », *L'Année sociologique*, vol. 47, n° 1, 1997, pp. 37-54.
- Ficara Annamaria, Cavallaro Lucia, Curreri Francesco, Fiumara Giacomo, De Meo Pasquale, Bagdasar Ovidiu, Song Wei et Liotta Antonio, « Criminal Networks Analysis in Missing Data scenarios through Graph Distances », *PLOS ONE*, vol. 16, n° 8, 2021, pp. 1-18.
- De Figueiredo Miguel, *Évaluation d'une approche chimiométrique non ciblée pour l'inférence de source de liquides inflammables en science forensique*, Université Paris-Saclay, Paris, 2018, 259 p.
- Fijnaut C. et Van Ostrive Lode, « Recherches sur la police », *Déviance et société*, vol. 2, n° 2, 1978, pp. 215-231.

Forget Amélie, « Du réseau comme structure au réseau comme acteur : comment les réseaux sociaux se constituent en acteurs collectifs », dans *Les réseaux criminels*, Presses universitaires de Montréal., Montréal, 2016, pp. 83-101.

Forsé Michel, « Les réseaux sociaux d'aujourd'hui : Un monde décidément bien petit », *Revue de l'OFCE*, vol. 126, n° 7, 2012, pp. 155.

Freeman Linton C., « Turning a profit from mathematics: The case of social networks », *The Journal of Mathematical Sociology*, vol. 10, n° 3-4, 12/1984, pp. 343-360.

Freeman Linton C. , « Centrality in social networks conceptual clarification », *Social Networks*, vol. 1, n° 3, 1978, pp. 215-239.

Gay Brigitte, Dousset Bernard et Wanassi Radwen, « Les indicateurs structurels d'un graphe : calculs, visualisation, interactivité », 5ème séminaire de Veille Stratégique Scientifique & Technologique (VSST 2013), Octobre 2013, Nancy, France.

Gayraud Jean-François et Farcy François, *Le renseignement criminel*, Paris, CNRS éd, 2014, 136 p.

Ghernaoui-Hélie Solange, *La cybercriminalité : le visible et l'invisible*, Lausanne, Presses polytechniques et universitaires romandes, 2009, 124 p.

Girod Alexandre, Champod Christophe et Ribaux Olivier, *Traces de souliers*, Lausanne, Presses polytechniques et universitaires romandes, 2008.

Gottfredson Michael R. et Hirschi Travis, *A general theory of crime*, Stanford, Calif, Stanford University Press, 1990, 297 p.

Hage Per et Harary Frank, *Structural models in anthropology*, Cambridge [Cambridgeshire] ; New York, Cambridge University Press, 1983, 201 p.

Harary Frank, *Graph Theory*, New York, Addison-Wesley Publishing Company, 1969, 286 p.

Harary Frank et Norman Robert Z., *Graph Theory as a Mathematical Model in Social Science*, Ann Arbor, University of Michigan, Institute for Social Research, 1953, 45 p.

Hoaglin David C., Mosteller Frederick et Tukey John Wilder (dir.), *Exploring data tables, trends, and shapes*, New York, Wiley, 1985, 527 p.

Holstrom John, « Vollmer as a man: Memories of a Close Friend and Colleague », dans *August Vollmer: Pioneer in Police Professionalism: Oral History Transcript/ and Related Material*, R.O.H. Office., Berkeley, University of California, 1972.

Honneth Axel et Rusch Pierre, *La lutte pour la reconnaissance*, Paris, Gallimard, 2015, 352 p.

Icduygu Ahmet et Toktas Sule, « How do Smuggling and Trafficking Operate via Irregular Border Crossings in the Middle East? Evidence from Fieldwork in Turkey », *International Migration*, vol. 40, n° 6, 12/2002, pp. 25-54.

Interpol (Groupe de travail sur l'analyse criminelle), *Guide sur l'analyse criminelle*, Lyon, Interpol, 1997, 66 p.

Jensen Pablo, *Pourquoi la société ne se laisse pas mettre en équations*, Paris, Éditions du seuil, 2018, 314 p.

Jourdain Anne et Naulin Sidonie, *La sociologie de Pierre Bourdieu*, Malakoff, Armand Colin, 2019, 672 p.

Kalifa Dominique, *Crime et culture au XIXe siècle*, Paris, Perrin, 2005, 331 p.

Katz Jack, *Seductions of crime: moral and sensual attractions in doing evil*, New York, Basic Books, 1988, 367 p.

Kempf Olivier, « Guerre informationnelle et jeux d'influence dans le cyberspace : le décryptage d'Olivier Kempf », *Communication & influence*, vol. 92, n° 1, 2018, pp. 6.

Keneally Thomas, *The commonwealth of thieves: the Sydney experiment*, Milsons Point, NSW, Random House Australia, 2005, 509 p.

Kirk Paul Leland, « Standardization of Criminological Nomenclature », *Journal of Criminal Law & Criminology*, vol. 38, n° 2, 1947, pp. 165-167.

Klerks Peter, « The Network Paradigm Applied to Criminal Organisations: Theoretical nitpicking or a relevant doctrine for investigators? Recent developments in the Netherlands », *Connections*, vol. 24, n° 1, 2001, pp. 53-65.

Knoke David et Burt Ronald S., « Prominence », dans Burt Ronald S. et Minor M. J. (dir.), *Applied Network Analysis*, Newbury Park, CA: Sage, 1983, pp. 195-222.

Lacroix Michel, « Littérature, analyse de réseaux et centralité : esquisse d'une théorisation du lien social concret en littérature », *Recherches sociographiques*, vol. 44, n° 3, 04/05/2004, pp. 475-497.

Lannier Salomé, « Blanchiment de la traite des êtres humains : réflexions générales sur la complémentarité des deux approches », *Cahiers de la sécurité et de la justice*, n° 48-49, 2020, pp. 271.

Latapy Matthieu, Magnien Clémence et Del Vecchio Nathalie, « Basic notions for the analysis of large two-mode networks », *Social Networks*, vol. 30, n° 1, 2008, pp. 31-48.

Laurent Sébastien-Yves et Fondation nationale des sciences politiques (dir.), *Archives « secrètes, » secrets d'archives ? L'historien et l'archiviste face aux archives sensibles*, Paris, CNRS, 2003, 288 p.

Laurent Sébastien-Yves, *Conflicts, Crimes and Regulations in Cyberspace*, London-Oboken, Iste-Wiley, 2021, 203 p.

Laurent Sébastien-Yves, *La conflictualité armée : approches interdisciplinaires*, Paris, Pedone, 2021, 276 p.

Laurent Sébastien-Yves, « Ce que le Cyber (ne) fait (pas) aux Relations internationales », *Etudes internationales*, LI, n° 2, t/2020, pp. 209-233.

Laurent Sébastien-Yves, « L'analyse, la ressource cachée du renseignement », *Revue des deux mondes*, 2016, pp. 60-66.

Laurent Sébastien-Yves, *Atlas du renseignement. Géopolitique du pouvoir*, Paris, Les Presses de science Po, 2014, 190 p.

Laurent Sébastien-Yves, « Plaidoyer pour un tournant analytique du renseignement et de la prospective », *Revue nationale défense*, n° 775, 2014, pp. 15-22.

Laurent Sébastien-Yves, *Daniel Halévy (1872-1962) : une écriture entre littérature et politique : du libéralisme au traditionalisme*, These de doctorat, Paris, Institut d'études politiques, 1999.

Laurent Sébastien-Yves, Forcade Olivier et Warusfel Bertrand, « Archives et politiques du secret : Ramener l'État à la raison démocratique », *Esprit*, n° 475, 2021, pp. 99-109.

Lavaud-Legendre Bénédicte, Melançon Guy, Pinaud Bruno, Plessard Cécile et Feron Norbert, « Analyse et visualisation des réseaux criminels », pp. 125.

Le Moigne Jean Louis, *La modélisation des systèmes complexes*, Paris, Dunod, 2007, 178 p.

Lehoux-Richer Magali, *Déstabilisation d'un réseau criminel pratiquant le trafic de stupéfiants*, Université de Montréal, 2009, 131 p.

Lemercier Claire, « Analyse de réseaux et histoire », *Revue d'histoire moderne & contemporaine*, vol. 52, n° 2, 2005, pp. 88-112.

Lenoir Rémi, « Espace social et classes sociales chez Pierre Bourdieu », *Sociétés & représentations*, vol. 17, n° 1, 2004, pp. 385-396.

Susan A. Lentz et Robert H. Chaires, « The invention of Peel's principles: A study of policing 'textbook' history », *Journal of Criminal Justice*, vol. 35, n° 1, 2007, pp. 69-79.

Leroux Luc, « Un tentaculaire réseau de blanchiment d'argent en procès à Marseille », *Le Monde*, 29/05/2019, p.

Letskemann Paul, *Crime as work*, Englewood Cliffs, 1973, 182 p.

Levitt Steven D. et Venkatesh Sudhir Alladi, « An Economic Analysis of a Drug-Selling Gang's Finances* », *Quarterly Journal of Economics*, vol. 115, n° 3, 2000, pp. 755-789.

Lindsay Murray R., « Publication System Biases Associated with the Statistical Testing Paradigm* », *Contemporary Accounting Research*, vol. 11, n° 1, 1994, pp. 33-57.

Lusher Dean, Koskinen Johan et Robins Garry (dir.), *Exponential random graph models for social networks: theory, methods, and applications*, Cambridge, Cambridge University Press, 2013, 336 p.

Malandain Gilles, « Les mouches de la police et le vol des mots : Les gazetins de la police secrète et la surveillance de l'expression publique à Paris au deuxième quart du XVIIIe siècle », *Revue d'histoire moderne et contemporaine*, vol. 42, n° 3, 09/1995, pp. 376-404.

Matloff Norman S., *The art of R programming: tour of statistical software design*, San Francisco, No Starch Press, 2011, 373 p.

Matsueda Ross L., Gartner Rosemary, Piliavin Irving et Polakowski, Michael « The Prestige of Criminal and Conventional Occupations: A Subcultural Model of Criminal Activity », *American Sociological Review*, vol. 57, n° 6, 1992, pp. 752.

Maurice Stéphanie, « Les liaisons dangereuses entre la Gendarmerie et un trafiquant d'armes », *Libération*, 2017, p.

McCarthy Bill et Hagan John, « When Crime Pays: Capital, Competence, and Criminal Success », *Social Forces*, vol. 79, n° 3, 2001, pp. 1035-1060.

McGloin J. M., « Policy Intervention Considerations of a Network Analysis of Street Gangs », *Criminology & Public Policy*, vol. 4, n° 1, 2005, pp. 607-635.

McNair Arnold D., « A History of English Criminal Law and Its Administration », *The Cambridge Law Journal*, vol. 10, n° 2, 1949, pp. 208-212.

Mercklé Pierre, *Sociologie des réseaux sociaux*, Paris, Découverte, 2004, 121 p.

Merton Robert K., « Social Structure and Anomie », *American Sociological Review*, vol. 3, n° 6, 1938, pp. 672-678.

Mitsilegas Valsamis, « From National to Global, from Empirical to Legal: The Ambivalent Concept of Transnational Organized Crime », dans Margaret E. Beare (dir.), *Critical Reflections on Transnational Organized Crime, Money Laundering, and Corruption*, Toronto, University of Toronto Press, 2003, pp. 55-87.

Mokhtari Karim et Carle Charlie, *Rédemption : itinéraire d'un enfant cassé*, Paris, Scrineo, 2013, 448 p.

Moreno J. L. et Jennings H. H., « Statistics of Social Configurations », *Sociometry*, vol. 1, n° 3/4, 1938, pp. 342-374.

Moreno Jacob Levy, « Sociogram and Sociomatrix: A Note to the Paper by Forsyth and Katz », *Sociometry*, vol. 9, n° 4, 1946, pp. 348-349.

Moreno Jacob Levy, *Who Shall Survive: A New Approach to the Problem of Human Interrelations*, Washington, DC, Nervous and Mental Disease Publishing Co., 1934, 460 p.

Morgan Anthony et Payne Jason, « Organised crime and criminal careers: Findings from an Australian sample. », *Trends & Issues in Crime & Criminal Justice*, n° 637, 2021, pp. 1-20.

Morselli Carlo, « Assessing Vulnerable and Strategic Positions in a Criminal Network », *Journal of Contemporary Criminal Justice*, vol. 26, n° 4, 2010, pp. 382-392.

Morselli Carlo, *Inside criminal networks*, New York, NY, Springer, 2009, 203 p.

Morselli Carlo et Roy Julie, « Brokerage qualifications in ringing operations », *Criminology*, vol. 46, n° 1, 2008, pp. 71-98.

Morselli Carlo et Tremblay Pierre, « Délinquance, performance et capital social : une théorie sociologique des carrières criminelles », *Criminologie*, vol. 37, n° 2, 2005, pp. 89-122.

Morselli Carlo et Tremblay Pierre, « Criminal achievement, offender networks and the benefits of low self-control », *Criminology*, vol. 42, n° 3, 08/2004, pp. 773-804.

Morselli Carlo et Tremblay Pierre, « Patterns in criminal achievement: Wilson and Abrahamse revisited », *Criminology*, vol. 38, n° 2, 1992, pp. 633-659.

Morselli Carlo, Tremblay Pierre et McCarthy Bill, « Mentors and criminal achievement », *Criminology*, vol. 44, n° 1, 2006, pp. 17-43.

Morselli Carlo, Turcotte Mathilde et Louis Guillaume, « Le crime organisé et les contre-mesures », dans *Traité de sécurité intérieure*, Lausanne, Presses polytechniques et universitaires romandes, 2007, pp. 185-199.

Mucchielli Laurent, *L'invention de la violence : des peurs, des chiffres et des faits*, Paris, Fayard, 2011, 339 p.

Mugg Quentin et Constanty Hélène, *Argent sale : la traque*, Paris, Fayard, 2021.

Notin Jean-Christophe, *Les guerriers de l'ombre*, 2017, 304 p.

Ogien Albert, *Sociologie de la déviance*, 2018, 261 p.

Oliver Willard M., *August Vollmer: the father of American policing*, Durham, North Carolina, Carolina Academic Press, 2017, 785 p.

Office des Nations Unies contre les Drogues et le Crime (ONUDDC), *Référentiel stratégique de lutte contre la criminalité organisée pour l'élaboration de stratégies à fort impact*, Vienne, Nations Unies, 2021.

Opsahl Tore, « Triadic closure in two-mode networks: Redefining the global and local clustering coefficients », *Social Networks*, vol. 35, n° 2, 2013, pp. 159-167.

- Paoli Letizia, *The Oxford handbook of organized crime*, Oxford, Oxford University Press, 2014, 692 p.
- Papaux Alain, « Les ressorts logiques de l'enquête: "l'erreur Sherlock Holmes" ; ou les fausses déductions d'un détective peu "rusé" », dans *Criminologie, politique criminelle et droit pénal dans une perspective internationale : mélanges en l'honneur de Martin Killias à l'occasion de son 65e anniversaire*, Berne, Stämpfli, 2013, pp. 1205-1220.
- Park Robert Ezra, Burgess Ernest Watson et McKenzie Roderick D., *The city: suggestions for investigation of human behavior in the urban environment*, Chicago, Univ. of Chicago Press, 2010, 239 p.
- Pearson Geoffrey et Hobbs Dick, *Middle Market Drug Distribution: Home Office Research Study*, Home Office, 2001, 91 p.
- Perliger Arie et Pedahzur, Ami « Social Network Analysis in the Study of Terrorism and Political Violence », *Political Science & Politics*, vol. 44, n° 01, 2011, pp. 45-50.
- Perrot Patrick, « L'analyse du risque criminel : l'émergence d'une nouvelle approche », *Revue de l'électricité et de l'électronique*, vol. 8, , 2014, pp. 100-107.
- Perrot Patrick, « Le renseignement criminel: de nouvelles perspectives contre la criminalité organisée », *Revue nationale défense*, vol. 4, n° 779, 2015, pp. 11-15.
- Peterson Marilyn B., *Applications in criminal analysis: a sourcebook*, Westport, Conn, Greenwood Press, 1994, 312 p.
- Piliavin Irving, Gartner Rosemary, Thornton Craig et Matsueda Ross L., « Crime, Deterrence, and Rational Choice », *American Sociological Review*, vol. 51, n° 1, 1986, pp. 101.
- Pinaud Bruno, Benichou Maud et Melançon Guy, « Extraction d'un réseau social criminel par transformation d'un graphe d'enquête multivarié », Conférence Extraction et Gestion des Connaissances (EGC), Lyon, 2023.
- Plotinus, Hadot Pierre, et Plotinus, *Traité 38: VI, 7*, Paris, Éditions du Cerf, 1988, 427 p.
- Polsky Ned, *Hustlers, beats and others*, Chicago, 1967, 218 p.
- Radzinowicz Leon, *A history of English criminal law and its administration from 1750*, London, Stevens, 1948.

- Rainer Robert, « La tradition policière britannique : modèle ou mythe ? », *Cahiers de la sécurité intérieure*, n° 7, 1992/1991, pp. 27-39.
- Ramos Daniel, Gonzalez-Rodriguez Joaquin, Zadora Grzegorz et Aitken Colin, « Information-Theoretical Assessment of the Performance of Likelihood Ratio Computation Methods », *Journal of Forensic Sciences*, vol. 58, n° 6, 2013, pp. 1503-1518.
- Ratcliffe Jerry H., *Intelligence-led policing*, London ; New York, Routledge, Taylor & Francis Group, 2016, 222 p.
- Ratcliffe Jerry H. (dir.), *Strategic thinking in criminal intelligence*, Sydney, Federation Press, 2009, 273 p.
- Ratcliffe Jerry H., « Intelligence-led policing: Anticipating risk and influencing action », dans *Criminal intelligence for the 21st century*, IALEIA, 2011, pp. 206-220.
- Regional Oral History Office, *August Vollmer: Pioneer in Police Professionalism: Oral History Transcript/ and Related Material*, Berkeley, University of California, 1972.
- Ribaux Olivier, *Police scientifique : le renseignement par la trace*, Lausanne, Presses polytechniques et universitaires romandes, 2014.
- Robitaille Clément, « À qui profite le crime ? Les facteurs individuels de la réussite criminelle », *Criminologie*, vol. 37, n° 2, 2005, pp. 33-62.
- Rossi Peter H., « Intrasocietal Consensus in Occupational Prestige Evaluations », dans *Occupational Prestige in Comparative Perspective.*, Saint Louis, Elsevier Science, 2013, pp. 59-78.
- Rossy Quentin, « La visualisation relationnelle au service de l'enquête criminelle », dans *Les réseaux criminels*, Les Presses de l'université de Montréal., Montréal, 2016, pp. 17-39.
- Sapiro Gisèle, Denord François, Duval Julien, Hauchecorne Mathieu, Heilbron Johan, Poupeau Franck et Seiler Hélène (dir.), *Dictionnaire international Bourdieu*, Paris, CNRS éditions, 2020, 963 p.
- Sellin Thorsten, « Organized Crime: A Business Enterprise », *The Annals of the American Academy of Political and Social Science*, vol. 347, , 1963, pp. 12-19.

Shaw Clifford R. et McKay Henry D., « Juvenile Delinquency and Urban Areas: A Study of Rates of Delinquents in Relation to Differential Characteristics of Local Communities in American Cities. », *American Journal of Sociology*, vol. 49, n° 1, 1943, pp. 100-101.

Sherman Lawrence W., Gartin Patrick R. et Buerger Michael E., « Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place », *Criminology*, vol. 27, n° 1, 02/1989, pp. 27-56.

Sikkink Kathryn, « The power of networks in international politics », dans *Networked politics: agency, power, and governance*, Ithaca, Cornell University Press, 2015, pp. 228-247.

Simmel Georg, Lilyane Deroche-Gurcel et Sibylle Muller, *Sociologie : étude sur les formes de la socialisation*, Paris, PUF, 2013, 768 p.

Sing Tobias, Sander Oliver, Beerenwinkel Niko, Lengauer Thomas, Unterthiner Thomas et Ernst Felix G. M., « ROCR: Visualizing the Performance of Scoring Classifiers ».

Skillicorn D. B. et Leuprecht C., « Copresence networks: criminals in their environment », *Social Network Analysis and Mining*, vol. 11, n° 1, 25/09/2021, pp. 85.

Skillicorn David B., Zheng Q. et Morselli Carlo, « Spectral embedding for dynamic social networks », Niagara, Ontario, Canada, ACM Press, 2013.

Soto Mauricio Abel Gomez, « Quelques propriétés topologiques des graphes et applications à Internet et aux réseaux », *Réseaux et télécommunications*, Paris Diderot University, 2011, pp. 130.

Sparrow Malcolm K., « The application of network analysis to criminal intelligence: An assessment of the prospects », *Social Networks*, vol. 13, n° 3, 1991, pp. 251-274.

Sullivan Mercer L., « *Getting paid* »: *youth crime and work in the inner city*, Ithaca, Cornell University Press, 1989, 275 p.

Sutherland Edwin H., *Le voleur professionnel : d'après le récit d'un voleur de profession*, 1963, 162 p.

Sutherland Edwin H., *Principles of criminology*, New York, General Hall Inc., U.S., 1947, 704 p.

Tannenbaum Frank, *Crime and the Community*, Columbia, Columbia University Press, 1938, 112 p.

Tarissan Fabien Frédéric, *Au coeur des réseaux: des sciences aux citoyens*, Paris, Editions le Pommier, 2019, 160 p.

Taroni Franco, *Data analysis in forensic science: a Bayesian decision perspective*, Chichester, John Wiley and Sons, 2010, 367 p.

Tenti Valentina et Morselli Carlo, « Group co-offending networks in Italy's illegal drug trade », *Crime, Law and Social Change*, vol. 62, n° 1, 2014, pp. 21-44.

Thomas William Isaac, *On social organization and social personality*, Chicago, The University of Chicago Press, 1966, 311 p.

Thomas William Isaac et Znaniecki Florian, *Le paysan polonais en Europe et en Amérique : récit de vie d'un migrant*, Paris, A. Colin, 2005, 446 p.

Thrasher Frederic Milton, *The gang: a study of 1,313 gangs in Chicago*, Chicago, Illinois, The University of Chicago Press, 2013, 605 p.

Treiman Donald J, *Occupational prestige in comparative perspective*, New York, Academic Press, 1977, 536 p.

Tremblay Pierre, *Le délinquant idéal : performance, discipline, solidarité*, Montréal, QC, Liber, 2010, 263 p.

Tremblay Pierre, « Attrition, récidive et adaptation », *Revue internationale de criminologie et de police technique et scientifique*, vol. 52, n° 2, 1999, pp. 163-178.

Tubaro Paola, *International encyclopedia of the social & behavioral sciences*, Amsterdam, Elsevier, 2015, 26 p.

Tzu Sun, *L'art de la guerre : Mille et une nuits.*, Paris, Pocket, 2000, 128 p.

United States DoD, *Countering Threat Networks*, CreateSpace Independent Publishing Platform, 2016, 164 p.

Encyclopædia Universalis, *Jacob Levy Moreno*, <https://www.universalis.fr/encyclopedie/jacob-levy-moreno/>, consulté le 13 novembre 2022.

United Nations Office on Drugs and Crime (UNODC), « Criminal Intelligence: Manual for Analysts », *Publishing and Library Section, United Nations Office at Vienna*, 2011, pp. 104.

Vernier Éric, *Techniques de blanchiment et moyens de lutte*, Paris, Dunod, 2005, 304 p.

- Vold George B., Bernard Thomas J. et Snipes Jeffrey B., *Theoretical criminology*, New York, Oxford University Press, 2002, 433 p.
- Vollmer August, « The Scientific Policeman », *The American Journal of Police Science*, vol. 1, n° 1, 1930, pp. 8.
- Vollmer August, « Revision of the Atcherley Modus Operandi System », *Journal of the American Institute of Criminal Law and Criminology*, vol. 10, n° 2, 1919, pp. 229-274.
- Wasserman Stanley et Faust Katherine, *Social Network Analysis Methods and Applications*, Cambridge, Cambridge University Press, 1994, 826 p.
- Whyte William Foote, *Street corner society: la structure sociale d'un quartier italo-américain*, Paris, La Découverte, 1943, 406 p.
- Wilson James Q. et Abrahamse Allan, « Does crime pay? », *Justice Quarterly*, vol. 9, n° 3, 1992, pp. 359-377.
- Wilson O. W, *Police planning*, Springfield, Ill., Thomas, 1977.
- Wilson O. W, « August Vollmer », *The Journal of Criminal Law, Criminology, and Police Science*, vol. 44, n° 1, 05/1953, pp. 91.
- Wilson O. W et McLaren Roy Clinton, *Police administration*, New York, McGraw-Hill, 1972, 742 p.
- Wilson O. W, *Police Records Their Installation and Use*, Berkeley, Public Administration Service, 1951, 376 p.
- Wolfgang Marvin E., Figlio Robert M. et Sellin Thorsten, *Delinquency in a Birth Cohort*, University of Chicago Press, 1972, 116 p.
- Wolfgang Marvin E. et Tracy Paul E., « The 1945 and 1958 birth cohorts: a comparison of the prevalence, incidence, and severity of delinquent behavior », Philadelphia, Pa., Center for Studies in Criminology and Criminal Law, University of Pennsylvania, 1982.
- Le grand Larousse illustré*, Paris, Larousse, 2019.
- Cifre-Défense*, <https://www.anrt.asso.fr/fr/cifre-defense-24859>, consulté le 28 mai 2022.
- Le dispositif Cifre*, <https://www.anrt.asso.fr/fr/le-dispositif-cifre-7844>, consulté le 28 mai 2022.

Édit de Louis XIV portant création de la lieutenance générale de police, 1667.

August Vollmer : pioneer in police professionalism. oral history transcript / 1972-1983.,
<https://oac.cdlib.org/search?style=oac4;Institution=UC%20Berkeley::Bancroft%20Library;idT=UCb11234320x>, consulté le 12 mai 2022.

Code pénal - Section III - De la définition de certaines circonstances entraînant l'aggravation, la diminution ou l'exemption des peines (L. No 2004-204 du 9 mars 2004, art. 12-I.) Dalloz,
https://www.dalloz.fr/documentation/Document?id=CODE_CPEN_ARTI_132-71&scrl=CPEN122562, consulté le 4 octobre 2022.

Les CIFRE, <https://www.enseignementsup-recherche.gouv.fr/fr/les-cifre-46510>, consulté le 28 mai 2022.

Table des tableaux et figures

Tableaux :

Tableau 1	Synthèse de la théorie de l'étiquetage de Becker.....	- 58 -
Tableau 2	Interprétation de différentes mesures de centralité.....	- 72 -
Tableau 3	En-têtes de colonnes concernant l'entité source dans l'onglet « personnes »-	89 -
Tableau 4	En-têtes de colonnes concernant le lien, communes à tous les onglets.....	- 89 -
Tableau 5	En-têtes de colonnes concernant l'entité destination, communes à tous les onglets	- 89 -
Tableau 6	Liste initiale des différentes entités (information stockée dans la colonne « icônes »)	- 90 -
Tableau 7	Répartition des entités par type	- 90 -
Tableau 8	Colonnes concernant l'entité source dans l'onglet « personnes », exemple de données et taux de remplissage.....	- 91 -
Tableau 9	Colonnes concernant l'entité source dans l'onglet « personnes morales », exemple de données et taux de remplissage	- 91 -
Tableau 10	Colonnes concernant l'entité source dans l'onglet « télécom », exemple de données et taux de remplissage.....	- 91 -
Tableau 11	Colonnes concernant l'entité source dans l'onglet « lieux », exemple de données et taux de remplissage.....	- 91 -
Tableau 12	Colonnes concernant l'entité source (dans l'onglet « objets »), exemple de données et taux de remplissage.....	- 92 -
Tableau 13	Colonnes concernant l'entité source (dans l'onglet « véhicules »), exemple de	

données et taux de remplissage.....	- 92 -
Tableau 14 Taux de remplissage des attributs des entités dans la matrice, par onglet et au total sur l'ensemble des données.....	- 92 -
Tableau 15 Répartition des liens par type d'entité source	- 93 -
Tableau 16 Répartition des liens par type	- 94 -
Tableau 17 Volume de liens confirmés et hypothétiques.....	- 94 -
Tableau 18 Colonnes concernant les liens (communes à tous les onglets), exemple de données et taux de remplissage.....	- 95 -
Tableau 19 Structure et exemple de données après réunion des onglets.	- 104 -
Tableau 20 Création d'identifiants uniques pour chaque entité.....	- 105 -
Tableau 21 Homogénéisation des labels des entités	- 105 -
Tableau 22 Exemple d'anomalie sémantique	- 106 -
Tableau 23 Exemple de standardisation des données	- 106 -
Tableau 24 Liste initiale des différentes entités (information stockée dans la colonne « icônes »)	- 106 -
Tableau 25 Nomenclature établie pour référencer les entités par icônes.....	- 107 -
Tableau 26 Stockage initial de certains liens présumés	- 107 -
Tableau 27 Stockage après standardisation des liens présumés.....	- 108 -
Tableau 28 Ontologie des entités	- 109 -
Tableau 29 Liste des catégories d'entités utilisées dans la suite du travail	- 109 -
Tableau 30 Extrait de l'ontologie de classification sémantique des liens.....	- 110 -

Tableau 31	Liste des catégories de liens utilisées dans la suite du travail.....	- 111 -
Tableau 32	15 premières lignes du fichier « <i>Nodes</i> »	- 113 -
Tableau 33	15 premières lignes du fichier « <i>Edges</i> »	- 113 -
Tableau 34	Mise en évidence (information confirmée) d'identités multiples	- 113 -
Tableau 35	Illustration de la gestion des liens fixes et ponctuels	- 116 -
Tableau 36	Illustration du vocabulaire emprunté aux grammaires formelles.....	- 122 -
Tableau 37	Description de la hiérarchie de sous-graphes utilisée avec <i>Tulip</i> pour réduire la complexité du graphe <i>n-mode</i>	- 137 -
Tableau 38	Nombre d'arêtes des 4574 graphes irréductibles	- 141 -
Tableau 39	Échelle de qualification/quantification des symboles terminaux.....	- 146 -
Tableau 40	Liste, occurrences et fréquence des 27 symboles terminaux soumis au questionnaire	- 148 -
Tableau 41	Résultats des réponses à la première des sept questions soumises aux analystes criminels de la GN	- 149 -
Tableau 42	Groupes sémantiques constitués pour l'attribution des forces liantes.....	- 152 -
Tableau 43	Énumération des symboles terminaux classés dans les groupes 2 et 3	- 152 -
Tableau 44	Exemple de valeurs de force liante	- 155 -
Tableau 45	Classes de force liante des liens et équivalence en pondération	- 157 -
Tableau 46	Décision sémantique et pondération des 25 <i>patterns</i> de chemins étudiés..	- 165 -
Tableau 47	<i>Patterns</i> mettant en évidence la nécessité d'améliorer le modèle.....	- 166 -
Tableau 48	Catégorisation des entités comme « LIEUX PUBLICS » ou « PERSONNES	

Tableau 49 Correction des pondérations de symboles minimaux concernant des « LIEUX PUBLICS » - 168 -

Tableau 50 Correction des pondérations de symboles minimaux concernant des « PERSONNES MORALES »..... - 169 -

Tableau 51 Décision sémantique et pondérations des 25 *patterns* de chemins étudiés avant (pondération1) et après (pondération2) corrections - 171 -

Tableau 52 Décision sémantique et pondération des 25 *patterns* de chemins étudiés.. - 173 -

Tableau 53 Légende - 174 -

Tableau 54 Décompte des VP, VN, FP, FN pour S_1 - 175 -

Tableau 55 Matrice de confusion pour $S_1 = 0.1$ - 175 -

Tableau 56 Décompte des VP, VN, FP, FN pour S_2 - 175 -

Tableau 57 Matrice de confusion pour $S_2 = 0.06$ - 175 -

Tableau 58 Problématiques distinctes dans des logiques d'enquête et de renseignement - 178 -

Tableau 59 Calcul de différents indicateurs pour chaque valeur seuil évaluée - 180 -

Tableau 60 Ontologie basée sur la sémantique et sur l'implication potentielle des structures - 190 -

Tableau 61 Caractéristiques du graphe et de ses 9 composantes connexes - 198 -

Tableau 62 Nombre ***Cdegrévi*** et proportion ***C'degrévi*** de voisins directs des 20 nœuds ayant les degrés les plus élevés, classés par centralité décroissante - 201 -

Tableau 63 Indicateurs de position et de dispersion pour la centralité de degré..... - 203 -

Tableau 64	Centralité d'intermédiarité et centralité d'intermédiarité centralisée des 20 premiers nœuds, classés par centralité décroissante	- 210 -
Tableau 65	Indicateurs de position et de dispersion pour la centralité d'intermédiarité-	211 -
Tableau 66	Profils de centralités dans l'espace graphique degré / intermédiarité	- 224 -
Tableau 67	Profils de centralités dans l'espace graphique degré primaire / intermédiarité secondaire	- 226 -
Tableau 68	Liste des prévenus (identifiants d'anonymisation) pour chacun des quatre profils de centralité étudié	- 230 -
Tableau 69	Résumé des comparaisons entre les résultats analytiques et les conclusions de l'enquête.	- 231 -

Figures :

Figure 1	Édit de Louis XIV portant création de la lieutenance générale de police - 7 -
Figure 2	Modèle des « 4P » de Ratcliffe - 36 -
Figure 3	Modèle du <i>DIKI</i> de Ratcliffe : de la donnée au renseignement - 37 -
Figure 4	Le cycle du renseignement - 37 -
Figure 5	Processus de veille opérationnelle..... - 40 -
Figure 6	Schéma illustrant l'hypothèse des champs et espaces criminels..... - 42 -
Figure 7	Illustration d'un système de positions absolues (les fonctions dans le réseau) et de positions sociales (place des acteurs les uns par rapport aux autres) dans un réseau de trafic de stupéfiants. - 43 -
Figure 8	Illustration d'un système de positions absolues (les fonctions dans les réseaux) et de positions sociales (place des acteurs les uns par rapport aux autres) dans une chaîne d'approvisionnement et de distribution de produits stupéfiants. - 44 -
Figure 9	Déclinaison du capital selon Bourdieu..... - 45 -
Figure 10	L'espace social selon Pierre Bourdieu..... - 48 -
Figure 11	Apparition des trois grandes théories du passage à l'acte - 50 -
Figure 12	(a) Graphe G (non-orienté) d'ordre 6 ; (b) Graphe G' orienté d'ordre 6..... - 65 -
Figure 13	Matrice d'adjacence du graphe G en figure 12 (a) - 66 -
Figure 14	Illustration des définitions énoncées - 66 -
Figure 15	Continuum du réseau comme structure au réseau comme acteur - 77 -
Figure 16	Exemple de graphe <i>two-mode</i> G représentant 11 individus et les événements auxquels ils ont assisté les 6 derniers mois. - 79 -

Figure 17	Réseau social <i>N</i> extrait du graphe G de la figure 16.	- 80 -
Figure 18	Modèle de la hiérarchie de graphes de <i>Tulip</i>	- 99 -
Figure 19	Le cycle du renseignement	- 100 -
Figure 20	Une donnée, deux besoins distincts.....	- 101 -
Figure 21	Sélection des attributs conservés	- 104 -
Figure 22	Illustration d'un méta-nœud (icône blanc) constitué de deux identités distinctes (icône rouges) appartenant à la même personne	- 114 -
Figure 23	Graphe relationnel, redessiné avec l'algorithme FM ³	- 114 -
Figure 24	Processus de détection et correction des anomalies dans les données	- 116 -
Figure 25	Extraits zoomés du graphe <i>n-mode</i> nettoyé.....	- 118 -
Figure 26	Exemples de chemins reliant graphiquement deux personnes (libellés initiaux) - 121 -	
Figure 27	Changement d'échelle sémantique pour utiliser les catégories de liens et d'entités (passage des libellés aux catégories).....	- 121 -
Figure 28	Analogie explicative de la notion de force liante	- 124 -
Figure 29	Illustration de l'indépendance de la sémantique et de la longueur des chemins. - 126 -	
Figure 30	Les nœuds de type « PERSONNE » ne peuvent être des nœuds intermédiaires - 127 -	
Figure 31	Triade interdite de Granovetter	- 127 -
Figure 32	Schéma de la structure en sous-graphes utilisée avec <i>Tulip</i> pour réduire la complexité du graphe.....	- 129 -

Figure 33	Principe de fusion de liens multiples.....	- 129 -
Figure 34	Retrait des nœuds et arêtes non pertinents (en jaune)	- 130 -
Figure 35	Fusion des liens multiples : situation initiale	- 131 -
Figure 36	Schéma de la méthode de fusion des arêtes en dérivation	- 132 -
Figure 37	Schéma de la méthode d'agrégation des arêtes en série en méta-arêtes	- 133 -
Figure 38	Conditions d'applications de l'agrégation en série (1).....	- 133 -
Figure 39	Conditions d'application de l'agrégation en série (2)	- 134 -
Figure 40	Conditions d'application de l'agrégation en série (3)	- 134 -
Figure 41	Informations contenues dans le chemin C	- 134 -
Figure 42	Informations contenues dans le chemin C'	- 135 -
Figure 43	Informations contenues dans le lien L	- 135 -
Figure 44	Processus itératif de fusion des arêtes en dérivation et d'agrégation des arêtes en série	- 136 -
Figure 45	Graphe $data_0$ (a) et graphe réduit $data_9$ (b).....	- 139 -
Figure 46	Exemple de chemin irréductible : les voitures et la carte SIM sont de degré 3, et il n'y a pas d'arêtes multiples.	- 140 -
Figure 47	Extrait zoomé de $data_9$	- 140 -
Figure 48	Exemple de sous-graphe irréductible	- 141 -
Figure 49	Réseau pondéré des personnes graphiquement connectées.....	- 144 -
Figure 50	Réseau pondéré des personnes graphiquement connectées (zoom).....	- 144 -

Figure 51	Illustration de la première des sept questions soumises aux analystes criminels de la GN	- 146 -
Figure 52	Distribution des valeurs majoritaires des réponses sur l'ensemble des 27 symboles terminaux	- 150 -
Figure 53	Distribution des moyennes des valeurs-réponses sur l'ensemble des 27 symboles terminaux soumis	- 150 -
Figure 54	Distribution des médianes des valeurs-réponses sur l'ensemble des 27 symboles terminaux soumis	- 150 -
Figure 55	Schéma de résumé de la méthode de pondération des liens directs entre individus	- 159 -
Figure 56	Graphes <i>data0</i> (à gauche) et <i>Social_network</i> (à droite). Les personnes sont représentées par les points bleus.	- 160 -
Figure 57	Zoom sur <i>Social_network</i> avec affichage des pondérations.....	- 160 -
Figure 58	Deux niveaux de zoom sur <i>Social_network</i> avec une échelle de couleur sur les pondérations. Les liens les plus clairs ont les plus faibles pondérations, les plus foncés ont les plus élevées.	- 161 -
Figure 59	Distribution des pondérations des arêtes de <i>Social_network</i> (échelle x en \log_{10})	- 163 -
Figure 60	Nécessité de prendre en compte pour certains liens, tels que ceux de type « localisation », de la date et de l'heure.....	- 170 -
Figure 61	Distribution des pondérations des arêtes de <i>Social_network</i> (échelle x en \log_{10}) après corrections	- 172 -
Figure 62	Distribution (a) et densité (b) des valeurs de pondération sur l'échantillon-	172
		-
Figure 63	Courbe ROC illustrant le TVP en fonction du TFN pour les différentes valeurs	

seuils (en vert).....	- 176 -
Figure 64 Graphe <i>Investigation_network</i> _{enq}	- 181 -
Figure 65 Graphe <i>Investigation_network</i> _{rens}	- 182 -
Figure 66 Position des seuils sur la courbe ROC	- 182 -
Figure 67 Stockage des entités difficilement classifiables dans des méta-nœuds	- 186 -
Figure 68 Calcul de la pondération du tronçon de chemin <i>T</i>	- 187 -
Figure 69 Illustration de la démarche suggérée pour contrôler les chaînes téléphoniques longues	- 187 -
Figure 70 Prise en compte de la disponibilité et des informations de dates et heures	- 188 -
Figure 71 Illustration de la démarche suggérée pour gérer les dates et heures sur les liens de type « localisation »	- 188 -
Figure 72 Prise en compte de l'avis expérimenté de l'enquêteur	- 190 -
Figure 73 Exemple de situation problématique	- 191 -
Figure 74 Sociogramme <i>INR</i>	- 192 -
Figure 75 Sociogramme <i>INR</i>	- 193 -
Figure 76 Sociogramme <i>INR</i> pré-traité.....	- 195 -
Figure 77 Le nœud V_5 est un cutpoint et l'arête e_{11} est un bridge.....	- 196 -
Figure 78 Exemple d'un ensemble de deux nœuds dont la suppression simultanée entraîne une fragmentation de <i>S</i>	- 197 -
Figure 79 Pondération de la taille des nœuds par leur centralité de degré. Les nœuds bleus sont les 20 acteurs ayant les plus forts degrés.....	- 200 -

Figure 80	Répartition des valeurs de centralité de degré (standardisée). (a) <i>Boxplot</i> ; (b) <i>Boxplot</i> sans les <i>outliers</i> ; (c) Courbe de densité	- 202 -
Figure 81	Les nœuds de la composante verte sont à une distance infinie des nœuds de la composante bleue (et réciproquement)	- 205 -
Figure 82	Pondération de la taille des nœuds par leur centralité de proximité (non standardisée). Les nœuds bleus sont les 20 acteurs ayant les plus fortes proximités.	- 206 -
Figure 83	Pondération de la taille des nœuds par leur centralité de proximité standardisée. Les nœuds bleus sont les 20 acteurs ayant les plus fortes proximités.	- 207 -
Figure 84	Pondération de la taille des nœuds par leur centralité d'intermédiation standardisée. Les nœuds orange sont les 20 acteurs ayant les plus fortes intermédiation.	- 209 -
Figure 85	Répartition des valeurs de centralité d'intermédiation (standardisée). (a) <i>Boxplot</i> ; (b) <i>Boxplot</i> sans les <i>outliers</i> ; (c) Courbe de densité	- 211 -
Figure 86	Exemple de position importante non détectée par les mesures de centralité primaires	- 213 -
Figure 87	Exemple de faible centralité primaire de degré mais de fort degré secondaire...	- 214 -
Figure 88	Pondération de la taille des nœuds par leur centralité « primaire » de degré. Les acteurs bleus sont les 20 acteurs ayant les plus forts degrés, préalablement mis en évidence. Les acteurs verts sont les 10 acteurs ayant les plus forts degrés secondaire (après exclusion de ceux appartenant déjà au premier groupe, au nombre de 14).....	- 215 -
Figure 89	Comparaison des <i>boxplot</i> de répartition des valeurs de degrés primaires et secondaires, avec les valeurs extrêmes (a) et sans (b).	- 216 -
Figure 90	<i>Scatterplot</i> représentant, pour chaque acteur, ses centralités primaires et secondaires de degré. L'équation de régression est : $y = 27.16924 \cdot x - 0.04498$	- 217 -
Figure 91	Schéma de la méthode des moindres carrés	- 218 -

Figure 92	Pondération de la taille des nœuds par leur centralité « primaire » de degré. Les acteurs oranges sont les 20 acteurs ayant les plus fortes intermédiarités, préalablement mis en évidence. Les acteurs violets sont les 10 acteurs ayant les plus fortes intermédiarités secondaires (après exclusion de ceux appartenant déjà au premier groupe).	- 220 -
Figure 93	Comparaison des <i>boxplot</i> de répartitions des valeurs de degrés primaires et secondaires, avec les valeurs extrêmes (a) et sans (b).	- 221 -
Figure 94	<i>Scatterplot</i> représentant, pour chaque acteur, ses centralités primaires et secondaires de degré.	- 221 -
Figure 95	Exemple de <i>broker</i> . L'acteur J est ici un intermédiaire essentiel entre les groupes bleu, vert et violet du réseau. Les différents acteurs dépendent donc de J pour accéder aux ressources et informations des groupes dont ils ne font pas partie.	- 223 -
Figure 96	<i>Scatterplot</i> représentant les positions des acteurs dans l'espace graphique des centralités de degré et d'intermédiarité.....	- 224 -
Figure 97	<i>Scatterplot</i> représentant les positions des acteurs dans l'espace graphique des centralités primaire de degré et secondaire d'intermédiarité.	- 225 -
Figure 98	<i>Scatterplot</i> représentant les positions des acteurs dans l'espace graphique des centralités de degré et d'intermédiarité.....	- 229 -
Figure 99	<i>Scatterplot</i> représentant les acteurs en fonction de leurs centralités primaire de degré et secondaire d'intermédiarité. Sont mis en évidence en rouge les prévenus désignés par leur identifiant d'anonymisation.	- 230 -
Figure 100	Continuum du réseau comme structure au réseau comme acteur	- 236 -
Figure 101	Courbes ROC avant et après correctifs pour un échantillon avec des longueurs de chemin de 2 à 5	- 326 -
Figure 102	Courbes ROC avant et après correctifs pour un échantillon avec des longueurs de chemin de 2 à 6	- 326 -

ANNEXES

Table des annexes

Annexe 1	Table des sigles et acronymes employés	- 278 -
Annexe 2	Peel's nine police principles	- 280 -
Annexe 3	Ontologie de classification sémantique des liens	- 281 -
Annexe 4	Circulaire ministérielle n°30053 du 3 septembre 2019 relative à l'export de données à destination de la recherche scientifique	- 302 -
Annexe 5	Liste, occurrences et fréquence des symboles terminaux existant dans les données prétraitées.....	- 307 -
Annexe 6	Questionnaire soumis aux analystes de la Gendarmerie nationale pour appuyer l'attribution des forces liantes des symboles terminaux	- 309 -
Annexe 7	Résultats du questionnaire.....	- 314 -
Annexe 8	Forces liantes des symboles terminaux (trinômes sémantiques) existants.-	324 -
Annexe 9	Comparaison des taux d'erreurs obtenus en considérant des <i>patterns</i> de chemins de longueur 2 à 5 et des <i>patterns</i> de longueurs 2 à 6 (avant et après correction)	- 326 -
Annexe 10	Détail des <i>patterns</i> de chemins évalués (échantillon) pour l'analyse des résultats de l'extraction du réseau social et la détermination de la valeur seuil de filtrage du réseau social d'enquête	- 328 -
Annexe 11	Détail des symboles minimaux dans l'échantillon et occurrences dans le fichier de données initial pré-traité.....	- 330 -
Annexe 12	Scripts <i>R</i> écrits avec l'interface <i>RStudio</i> pour les générer les graphiques d'analyse du modèle	- 331 -
Annexe 13	Interprétation des indicateurs de performance utilisés en complément des courbes ROC	- 333 -

Annexe 14	Scripts <i>R</i> écrits avec l'interface <i>RStudio</i> pour effectuer les visualisations et calculer les centralités secondaires	- 334 -
Annexe 15	Centralité de degré des 62 acteurs du réseau pour lesquels cet indicateur prend des valeurs extrêmes (par rapport à l'ensemble des 615 acteurs).....	- 342 -
Annexe 16	Centralité d'intermédiarité standardisée (pourcentage d'appartenance à l'un des chemins géodésiques existants entre deux nœuds quelconques) des 146 acteurs du réseau pour lesquels cet indicateur prend des valeurs extrêmes (par rapport à l'ensemble des 615 acteurs) -	343 -
Annexe 17	Liste des acteurs des quatre profils de centralité d'intérêt	- 347 -

Annexe 1 Table des sigles et acronymes employés

ACO : Analyse criminelle opérationnelle

AID : Agence de l'innovation de défense

ANRT : Association nationale de la recherche et de la technologie

ARS (*SNA*) : Analyse de réseaux sociaux (*Social Network Analysis*)

AUC: *Area Under Curve* (aire sous la courbe)

CIFRE : Conventions industrielles de formation par la recherche

CIR : Crédit d'impôt recherche

CR : Commission rogatoire

CRT : Commission rogatoire technique

DCO : Division de la criminalité organisée

DIKI: *Data – Information – Knowledge – Intelligence*.

DR : Division renseignement

DS : Division stratégique

EDMH : École doctorale Montaigne Humanités

ERGM : *Exponential Random Graph Models*

ESC : École des sciences criminelles (Université de Lausanne)

GCO : Groupe criminel organisé

GDV : Gens du voyage

GN : Gendarmerie nationale

GTAC : Groupe de travail sur l'analyse criminelle d'Interpol

HDR : Habilité.e à diriger des recherches

IAELIA: *International Association of Law Enforcement Intelligence Analysts*

ILP: *Intelligence-Led Policing*

IQR: *Inter-Quartile Range* (Écart interquartile)

IRCGN : Institut de recherche criminelle de la Gendarmerie nationale

IRM : Institut de recherche Montesquieu

LaBRI : Laboratoire bordelais de recherche en informatique

MCF : Maître de conférence

MESRI : Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation

NCIS : *National Crime Intelligence Service*

NIM : *National Intelligence Model*

PME : Petites et moyennes entreprises

PN : Police nationale

PTS : Police technique et scientifique

PV : Procès-verbal

ROEM : Renseignement d'origine électromagnétique

ROC (courbe) : *Receiver Operating Characteristic (curve)*

ROHUM : Renseignement d'origine humaine

ROIM : Renseignement d'origine image

SCRC (GN) : Service central de renseignement criminel (de la Gendarmerie nationale)

SGAT : Sécurité globale et analyse trilingue

SHS : Sciences humaines et sociales

SR : Section de recherche

TALN (*NLP*) : Traitement automatique du langage naturel (*Natural Language Processing*)

TFN : Taux de faux négatifs

TFP : Taux de faux positifs

TVN : Taux de vrais négatifs

TVP : Taux de vrais positifs

UBx : Université de Bordeaux

UBM : Université Bordeaux Montaigne

UPMC : Université Pierre et Marie Curie (Paris VI).

Annexe 2 “Peel’s nine police principles”⁴⁸⁹

1. To prevent crime and disorder, as an alternative to their repression by military force and severity of legal punishment.
2. To recognize always that the power of the police to fulfill their functions and duties is dependent on public approval of their existence, actions and behavior, and on their ability to secure and maintain public respect.
3. To recognize always that to secure and maintain the respect and approval of the public means also the securing of the willing cooperation of the public in the task of securing observance of laws.
4. To recognize always that the extent to which the cooperation of the public can be secured diminishes proportionately the necessity of the use of physical force and compulsion for achieving police objectives.
5. To seek and preserve public favor, not by pandering to public opinion, but by constantly demonstrating absolute impartial service to law, in complete independence of policy, and without regard to the justice or injustice of the substance of individual laws, by ready offering of individual service and friendship to all members of the public without regard to their wealth or social standing, by ready exercise of courtesy and friendly good humor, and by ready offering of individual sacrifice in protecting and preserving life.
6. To use physical force only when the exercise of persuasion, advice and warning is found to be insufficient to obtain public cooperation to an extent necessary to secure observance of law or to restore order, and to use only the minimum degree of physical force which is necessary on any particular occasion for achieving a police objective.
7. To maintain at all times a relationship with the public that gives reality to the historic tradition that the police are the public and that the public are the police, the police being only members of the public who are paid to give full-time attention to duties which are incumbent on every citizen in the interests of community welfare and existence.
8. To recognize always the need for strict adherence to police-executive functions, and to refrain from even seeming to usurp the powers of the judiciary of avenging individuals or the State, and of authoritatively judging guilt and punishing the guilty.
9. To recognize always that the test of police efficiency is the absence of crime and disorder, and not the visible evidence of police action in dealing with them.

⁴⁸⁹ Law Enforcement Action Partnership, *Sir Robert Peel’s Policing Principles*, <https://lawenforcementactionpartnership.org/peel-policing-principles/>, consulté le 6 mai 2022.

Annexe 3 Ontologie de classification sémantique des liens

Conformément aux accords de confidentialité signés avec le SCRC en vertu de la circulaire ministérielle n°30053 du 3 septembre 2019 relative à l'export de données à destination de la recherche scientifique.

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
1	Discute	En contact
2	Récupère 700000€	Échange objet ou argent
3	Conducteur	Utilisateur
4	Entre avec un sac en papier blanc	Échange objet ou argent
5	Est en contact (transaction)	En contact
6	Récupère un sac noir et blanc	Échange objet ou argent
7	Charge un sac noir et blanc dans	Échange objet ou argent
8	Remet un sac noir	Échange objet ou argent
9	Échange un sac vert	Échange objet ou argent
10	Discute avec	En contact
11	Monte cote passager	Utilisateur
12	Prend deux sacs dans	Échange objet ou argent
13	Est titulaire de l'abonnement edf depuis le 14/09/2016	Adresse
14	Donne à la police lors gav	Utilisateur
15	Est titulaire de l'abonnement edf du 23/01/2014 au 13/09/2016	Adresse
16	Se trouve	Localisation
17	Est arrêté par la police au volant	Utilisateur
18	Monte à bord après avoir chargé deux sacs	Utilisateur
19	Quitte l'aéroport de Roissy à bord	Utilisateur
20	Monte à bord et quitte le parking de l'appart 'hôtel torcy	Utilisateur
21	Quitte le 20 avenue de l'Idéal Bondy et quitte les lieux à bord	Utilisateur

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
22	Monte à bord et quitte les lieux	Utilisateur
23	Contact téléphonique au sujet Noémie avec son adresse et son numéro de téléphone (sono 20321)	En contact
24	Donne comme numéro lors réservation campanile bussy	Utilisateur
25	Passager	Utilisateur
26	Loue du 03/07/2016 au 23/07/2016	Utilisateur
27	Entre	Localisation
28	Descend	Utilisateur
29	Demeure (siv)	Adresse
30	A donne comme numéro lors ouverture compte nickel	Utilisateur
31	A donne comme adresse lors ouverture compte nickel	Adresse
32	Titulaire	Titulaire
33	Entre en contact	En contact
34	Est interpellé par la sûreté territoriale des hauts de seine avec	Antécédents
35	Utilise	Utilisateur
36	Mariée en 2013	Familial/amical
37	Demeure (passeport 02/03/2015)	Adresse
38	Envoie mandat cash	Échange objet ou argent
39	Est gérant de l'épicerie	Professionnel
40	Est propriétaire (n'est pas sur le pv)	Titulaire
41	A émis chèque (2013)	Échange objet ou argent
42	A fait remise de chèque dont émetteur (2014)	Échange objet ou argent
43	Est vraisemblablement	Est
44	A fait remise de chèque dont émetteur(2016)	Échange objet ou argent
45	Propriétaire	Titulaire
46	A émis chèque (2012)	Échange objet ou argent

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
47	A possède un contrat edf résilié en 1986	Échange objet ou argent
48	A vendu le véhicule à Malone Flora	Échange objet ou argent
49	Crédite un chèque de 4950€ le 11/07/2014	Échange objet ou argent
50	Quitte l'épicerie	Localisation
51	Utilise (fiche client cic)	Utilisateur
52	Épouse	Familial/amical
53	Demeure	Adresse
54	Fait entrer	En contact
55	Entretient relation amoureuse	Familial/amical
56	Se rend avec Blein Carine	Localisation
57	A possédé un contrat edf résilié en 1995	Adresse
58	Demeure (avril 2016)	Adresse
59	Se trouve pour embarquer vers malaga (vol ux1038)	Localisation
60	Demeure (décembre 2013)	Adresse
61	Utilise pour réservation vol malaga	Utilisateur
62	Passager avant	Utilisateur
63	Remet un sac au conducteur	Échange objet ou argent
64	Utilise identité pour la location de box a champigny	Utilisateur
65	Quitte les lieux à bord	Utilisateur
66	Dirigeant	Professionnel
67	Rencontre au campanile bussy saint georges	En contact
68	Utilisait (2013-2014 shugard)	Utilisateur
69	Utilise comme alias pour une location de box	Utilisateur
70	Possède un abonnement edf	Adresse
71	Remet un sac blanc	Échange objet ou argent

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
72	Demeure (agence location)	Adresse
73	Demeure (adresse donnée lors gav tabahriti bakhti)	Adresse
74	Loue du 17/06/2016 au 01/07/2016	Utilisateur
75	Est	Est
76	Utilise (2014 et 2016 - taj)	Utilisateur
77	Utilise (sono 20321)	Utilisateur
78	Récupère 1 sac a dos noir dans	Échange objet ou argent
79	Remet un sac	Échange objet ou argent
80	Acquéreur	Titulaire
81	Salue	En contact
82	Entre dans la résidence	Localisation
83	Inscrit sur la boite aux lettres	Adresse
84	Responsable taches administratives	Professionnel
85	Propriétaire d'un appartement	Titulaire
86	Marié par la loi coranique	Familial/amical
87	Discute avec conducteur	En contact
88	Entre dans le garage	Localisation
89	Se rend	Localisation
90	Dépose deux gros sacs dans le coffre	Échange objet ou argent
91	Demeure (2015)	Adresse
92	Demeure (2012)	Adresse
93	Demeure de 2002 a 2011	Adresse
94	Marié	Familial/amical
95	Utilise (juin 2015)	Utilisateur
96	Demeure (2004)	Adresse

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
97	Demeure (2014)	Adresse
98	Utilise (2014)	Utilisateur
99	Charge un sac en papier marron a l'arriere	Échange objet ou argent
100	Demeure (taj 2013, 2010, 2009, 2008, 2007)	Adresse
101	Utilise comme fausse identité	Utilisateur
102	Remet un sac de couleur noire	Échange objet ou argent
103	Utilise (shugard)	Utilisateur
104	Donne un sac de course	Échange objet ou argent
105	Se trouve à l'intérieur	Localisation
106	Adresse cabinet médical	Adresse
107	Demeure (adresse vérifiée)	Adresse
108	Remet un sac en papier blanc	Échange objet ou argent
109	Dépose le sac en papier blanc	Échange objet ou argent
110	Gérant	Professionnel
111	Employé (depuis 2015)	Professionnel
112	Utilise (2015)	Utilisateur
113	Demeure (fisc)	Adresse
114	Lien familial (numero nie lie au dni)	Familial/amical
115	Demeure (2013)	Adresse
116	Demeure (siv 2016)	Adresse
117	Demeure (2007)	Adresse
118	Gérante	Professionnel
119	Frère	Familial/amical
120	Sœur	Familial/amical
121	Demeure bat4, appt 2243	Adresse

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
122	Demeure (fnpc et agdref)	Adresse
123	Verbalise a 2 reprises au volant	Utilisateur
124	Employé du 11/05/2015 au 31/12/2015	Professionnel
125	Monte à bord	Utilisateur
126	Se rend avec bendah aomar	Localisation
127	Travaille	Professionnel
128	Quitte épicerie alimentation accompagné de Barev Alain	Localisation
129	Ouvre la porte et observe avant de refermer la porte	Localisation
130	Demeure appt 14	Adresse
131	Discute devant épicerie jourdain	En contact
132	Monte à bord et quitte l'épicerie jourdain	Utilisateur
133	A été commercial	Professionnel
134	Verbalisé en 2013 à bord	Utilisateur
135	Entre dans l'épicerie	Localisation
136	Ressort de l'épicerie avec un sac blanc	Échange objet ou argent
137	Utilise pour son activité	Utilisateur
138	Identifié comme	Est
139	Ressort	Localisation
140	A donné comme numéro pour abonnement edf	Utilisateur
141	Se promène sur les quais de seine avec	En contact
142	Paye abonnement edf	Adresse
143	Dépose un sac	Échange objet ou argent
144	Demeure (snpc)	Adresse
145	Fils	Familial/amical
146	Demeure (pc)	Adresse

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
147	A travaille du 01/01/2015 au 27/10/2015	Professionnel
148	Quitte l'épicerie avec un sac en cuir noir	Échange objet ou argent
149	Remet un sac en cuir noir	Échange objet ou argent
150	Est assis a la terrasse	Localisation
151	Sort de l'épicerie avec une pochette de couleur bleue	Échange objet ou argent
152	Passager arrière	Utilisateur
153	Demeure au 6eme étage, porte 560	Adresse
154	Loue	Utilisateur
155	Verbalisé le 03/08/2013 à bord	Utilisateur
156	Employé du 02/09/2014 au 31/12/2015	Professionnel
157	Employé du 01/12/2014 au 31/12/2015	Professionnel
158	Employé du 01/01/2015 au 31/12/2015	Professionnel
159	Employé du 03/11/2014 au 05/12/2014	Professionnel
160	Quitte les lieux avec un papier format a4	Échange objet ou argent
161	Se trouve dans l'agence	Localisation
162	Monte à bord et quitte les lieux (rue du mont cenis, paris 18)	Utilisateur
163	Monte dans le véhicule	Utilisateur
164	Titulaire depuis le 12/10/2012	Titulaire
165	Concubine	Familial/amical
166	Utilise (contrat edf 2012)	Utilisateur
167	Utilise (2012)	Utilisateur
168	Co-auteur infractions en 1997/1999	Criminel
169	Demeure (snpc/edf/gaz)	Adresse
170	Demeure (2011)	Adresse
171	Demeure (2016)	Adresse

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
172	Dépose le sac noir que daoudi abdelakrim lui a remis dans le coffre	Échange objet ou argent
173	Locataire	Utilisateur
174	Est identifié	Est
175	Titulaire et utilisateur	Titulaire
176	A récupéré 130 000€ entre le 31/03/2016 et le 29/07/2016 auprès	Échange objet ou argent
177	Gérant depuis 2009	Professionnel
178	En voiture avec	En contact
179	Sort	Localisation
180	Remet 2 sacs de sport	Échange objet ou argent
181	Ancien titulaire	Titulaire
182	Ancien gérant	Professionnel
183	Compagne	Familial/amical
184	Verbalise à bord	Utilisateur
185	Possède un contrat def depuis le 30/06/2016	Adresse
186	Demeure (contrat tel)	Adresse
187	Travaille depuis le 16/12/2013	Professionnel
188	Demeure (1992)	Adresse
189	Travaille depuis le 07/05/2014	Professionnel
190	Travaille depuis le 10/07/2014	Professionnel
191	Travaille depuis le 31/03/2014	Professionnel
192	Demeure (snpc 2014)	Adresse
193	Fille	Familial/amical
194	Titulaire depuis le 14/09/2016	Titulaire
195	Demeure (infogreffe)	Adresse
196	Effectue 13 407,78€ de virement	Échange objet ou argent

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
197	Père	Familial/amical
198	A des parts	Financier
199	Utilise (fisc)	Utilisateur
200	Est ou a été associé et dirigeant	Professionnel
201	Associé en 2004 pour achat bien au 127 rue de la présentation à paris 75011	En contact
202	Administrateur	Professionnel
203	Ancien administrateur (avant 2014)	Professionnel
204	Vendeur appartement	Autre
205	Titulaire depuis 13/09/2016	Titulaire
206	Demeure (pc, titre de séjour et cpam)	Adresse
207	Mis en cause (2005)	M.E.C.
208	Avec	En contact
209	Demeure (siv - ancienne adresse)	Adresse
210	Titulaire contrats erdf ouverts le 09/08 et le 10/08/2016	Adresse
211	En contact	En contact
212	Monte dans le véhicule et quitte l'aéroport	Utilisateur
213	Ami avec	En contact
214	Sms échangé	En contact
215	Monte dans le véhicule et quitte l'aéroport	Utilisateur
216	Se rend avec une femme et un enfant	Localisation
217	Gérant du 25/01/2013 au 22/12/2016	Professionnel
218	Rencontre	En contact
219	Entre avec xh1 perriers	Localisation
220	Entre dans la résidence avec	En contact
221	Monte à bord place conducteur	Utilisateur

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
222	Stationne devant	Localisation
223	Monte à bord place conducteur et quitte les lieux à bord	Utilisateur
224	Dépose un paquet	Échange objet ou argent
225	Se stationne à côté	Localisation
226	A travaille au sein des locaux	Professionnel
227	Ancien propriétaire	Autre
228	Demeure acte de vente du bien immobilier	Adresse
229	Demeure (enedis)	Adresse
230	Présidente	Professionnel
231	A récupéré 5 000 euros	Échange objet ou argent
232	A récupéré 10 000 euros	Échange objet ou argent
233	Se trouve dans agenda téléphone	En contact
234	A récupéré 30 000 euros	Échange objet ou argent
235	A récupéré 50 000 euros (entre le 30/10 et le 30/11/2016)	Échange objet ou argent
236	A récupéré 16 300 euros	Échange objet ou argent
237	A récupéré 12 000 euros (entre le 12 et le 19/05/2016)	Échange objet ou argent
238	Utilise (contrat enedis)	Utilisateur
239	A récupéré 10 000 euros (septembre 2016)	Échange objet ou argent
240	Envoi avis imposition chez	Adresse
241	Séparé en 2014 (caf)	Familial/amical
242	Utilise (caf)	Utilisateur
243	Mère	Familial/amical
244	Mère des enfants de	Familial/amical
245	Co-auteur pv 0975/2015/00045	Criminel
246	Demeure (agdref)	Adresse

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
247	Entre en contact avec	En contact
248	Entre au contact	En contact
249	Descend passager du vl	Utilisateur
250	Se stationne	Localisation
251	Monte à bord porteur d'un paquet blanc	Échange objet ou argent
252	Descend du véhicule	Utilisateur
253	Récupère carton sur la banquette arrière	Échange objet ou argent
254	En couple avec	Familial/amical
255	Demeure (erdf)	Adresse
256	Père des enfants de	Familial/amical
257	Demeure (taj)	Adresse
258	Co-auteur importation de stupéfiants (2014)	Criminel
259	Se trouve à bord place conducteur	Utilisateur
260	Monte place conducteur	Utilisateur
261	Monte à bord place passager avant et quitte les lieux à bord	Utilisateur
262	Se trouve à côté	Localisation
263	Entre dans la résidence avec xh1 ds-153-jz	Localisation
264	Récupère sac blanc	Échange objet ou argent
265	S'assoit place passager et quitte les lieux à bord	Utilisateur
266	Occupant	Utilisateur
267	Demeure (adoc)	Adresse
268	Co-auteur meurtre en bande organisée (2015)	Criminel
269	Co-auteur importation de stupéfiants (2016)	Criminel
270	Co-auteur blanchiment (2013)	Criminel
271	Co-titulaire	Titulaire

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
272	Monte place passager avant	Utilisateur
273	Occupe	Utilisateur
274	Conductrice	Utilisateur
275	Petite amie	Familial/amical
276	Se trouve au niveau	Localisation
277	Demeure (siv) chez Boual Katia, 2-ème étage, porte 204	Adresse
278	Demeure chez	Familial/amical
279	Demeure (contrat tel.)	Adresse
280	A récupéré 2 000 euros	Échange objet ou argent
281	A récupéré 3 500 euros	Échange objet ou argent
282	A récupéré 6 000 euros	Échange objet ou argent
283	A récupéré 3 000 euros	Échange objet ou argent
284	Dépose un casque	Échange objet ou argent
285	Utilise (taj)	Utilisateur
286	Demeure (siv - taj)	Adresse
287	Demeure (source : société darkot auto)	Adresse
288	Demeure (siv) chez Tirant Delphine	Adresse
289	Demeure (chez)	Adresse
290	Utilise (fnpc)	Utilisateur
291	Échange un paquet indéterminé	Échange objet ou argent
292	Même nom	Même nom
293	Fait partie	Fait partie
294	Monte sur le scooter et quitte les lieux	Utilisateur
295	Entre dans le bâtiment g	Localisation
296	Monte passager	Utilisateur

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
297	Se rend dans le parking souterrain	Localisation
298	Monte à l'arrière	Utilisateur
299	Entre dans	Localisation
300	Récupère quelque chose	Échange objet ou argent
301	Se stationne	Localisation
302	Sort de la résidence le Pierront accompagne de xf et 2 enfants	En contact
303	Se trouve à proximité	Localisation
304	Dépose une enveloppe kraft marron	Échange objet ou argent
305	Demeure (appartement 260, 4 -ème étage)	Adresse
306	Remet quelque chose à 2 mains	Échange objet ou argent
307	Récupèrent petits pochons sous le véhicule	Localisation
308	Discute longuement	En contact
309	Co-auteur trafic de stupéfiants	Criminel
310	Se trouve autour	Localisation
311	Utilisateur	Utilisateur
312	Monte sur le scooter	Utilisateur
313	Demeure (siv - aucun nom correspondant)	Adresse
314	A demeure	Adresse
315	Utilise (télépage savn - bip&go)	Utilisateur
316	Demeure (élépage savn - bip&go)	Adresse
317	Demeure (ficoba)	Adresse
318	Mariée	Familial/amical
319	Co-auteur contrefaçon	Criminel
320	Doit s'abstenir de recevoir ou rencontrer (cj)	Lien interdit
321	Se trouve devant	Localisation

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
322	Monte au 6 ou 7eme étage	Localisation
323	Consulte	Utilisateur
324	Demeure (sfr)	Adresse
325	Co-auteur meurtre, violences avec arme, dégradation vhl en 2002	Criminel
326	Co-auteur importation de stupéfiants en 2004	Criminel
327	Conducteur habituel	Utilisateur
328	A récupéré 25 000 euros	Échange objet ou argent
329	A récupéré 20 000 euros	Échange objet ou argent
330	A récupéré 50 000 euros	Échange objet ou argent
331	A récupéré 23 000 euros	Échange objet ou argent
332	A récupéré 30 000 euros (entre le 09/11 et le 15/11/2016)	Échange objet ou argent
333	A récupéré 20 000 euros (entre le 26 et 27/05/2016)	Échange objet ou argent
334	A récupéré 40 000 euros (entre le 15/06 et le 04/08/2016)	Échange objet ou argent
335	A récupéré 5 000 euros (entre le 09 et le 15/11/2016)	Échange objet ou argent
336	Associe	Professionnel
337	Marie (bcr)	Familial/amical
338	A récupéré 60 000 euros (entre le 01/09 et le 07/09/2016)	Échange objet ou argent
339	Demeure (fnpc)	Adresse
340	A récupéré 20 000 euros (entre le 15/06 et le 20/06/2016)	Échange objet ou argent
341	A récupéré 8 700 euros	Échange objet ou argent
342	Adresse fiscale chez	Adresse
343	Associée	Professionnel
344	A récupéré 2 500 euros	Échange objet ou argent
345	A récupéré 5 000 euros (entre le 03 et le 12/05/2016)	Échange objet ou argent
346	Utilise (enedis)	Utilisateur

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
347	Propriétaire (siv)	Titulaire
348	Travaille pour	Professionnel
349	Demeure (chez Bassar Yohan)	Adresse
350	Marie a	Familial/amical
351	Lien facebook	Familial/amical
352	Demeure (contrat enedis)	Adresse
353	Adresse d'imposition	Adresse
354	Directeur général	Professionnel
355	Gérant (radie depuis 2013)	Professionnel
356	Demeure (engie)	Adresse
357	Propriétaire (bcr)	Titulaire
358	A récupéré 25 000 euros (entre le 01 et le 07/09/2016)	Échange objet ou argent
359	Demeure (enedis)	Adresse
360	Demeure (bcr)	Adresse
361	Demeure au 3 -ème étage (enedis)	Adresse
362	Employé connu au 31/12/2016	Professionnel
363	Résidait en 2008	Adresse
364	Mariée (facebook)	Familial/amical
365	Propriétaire (03/12/2015)	Titulaire
366	Adresse profession libérale	Professionnel
367	Ex-mari	Familial/amical
368	Utilise (enedis edf / bcr)	Utilisateur
369	A récupéré 4 000 euros	Échange objet ou argent
370	A récupéré 52 000 euros	Échange objet ou argent
371	Appartient a	Titulaire

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
372	Amie (facebook)	Familial/amical
373	Mariée (état civil)	Familial/amical
374	A récupéré 35 000 euros (entre le 26/02 et le 09/03/2016)	Échange objet ou argent
375	Travaille avec	Professionnel
376	A récupéré 40 000 euros	Échange objet ou argent
377	Titulaire (enedis)	Titulaire
378	Divorce	Familial/amical
379	Titulaire (siv)	Titulaire
380	Employeur	Professionnel
381	Marié (état civil)	Familial/amical
382	Demeure (enedis edf)	Adresse
383	Inscription bal	Adresse
384	N'est pas le père de	Non lien
385	Directeur technique	Professionnel
386	Demeure (caf)	Adresse
387	Ami (facebook)	Familial/amical
388	Mariée (bcr)	Familial/amical
389	Dernière adresse d'imposition (bcr)	Adresse
390	Demeure (avis d'imposition - bcr)	Adresse
391	Demeure (probablement fausse)	Adresse
392	Titulaire (jusqu'au 21/02/2017)	Titulaire
393	Demeure (enedis, bcr)	Adresse
394	Pacse	Familial/amical
395	Employé (en 2016)	Professionnel
396	Demeure (contrat engie)	Adresse

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
397	Mariée (piac)	Familial/amical
398	Demeure (enedis résiliation 16/03/2017)	Adresse
399	Mariée (bcr)	Familial/amical
400	Propriétaire (bcr)	Titulaire
401	A récupéré 15 000 euros	Échange objet ou argent
402	A récupéré 20 000 euros (entre le 07 et 09/09/2016)	Échange objet ou argent
403	Se trouve dans le répertoire téléphone	En contact
404	Donne sac course vert	Échange objet ou argent
405	Sort du véhicule	Utilisateur
406	En contact avec utilisateur	Autre
407	À côté	Localisation
408	Se trouve à côté et fait le guet	Localisation
409	Prend bouteille de shampoing dans le coffre	Échange objet ou argent
410	Verbalisé à bord	Utilisateur
411	Verbalisé à bord	Utilisateur
412	Donne enveloppe blanche	Échange objet ou argent
413	Co-auteur association de malfaiteurs	Criminel
414	Demeure (opérateur)	Adresse
415	Titulaire (opérateur)	Titulaire
416	Transfert gros carton	Échange objet ou argent
417	Est en compagnie et l'appelle "William"	En contact
418	Sis	Adresse
419	Sise	Adresse
420	A un établissement	Adresse
421	A comme associé	Professionnel

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
422	Sise (siv)	Adresse
423	Situé près	Localisation
424	Située	Localisation
425	Situé	Localisation
426	Bailleur (bcr)	Titulaire
427	Possède contrat engie	Adresse
428	Ferme a cette adresse depuis 13/02/2016	Autre
429	Sise (caf)	Adresse
430	Titulaire (depuis 05/11/2013)	Titulaire
431	Syndic de copropriété	Professionnel
432	Contact téléphonique au sujet femme code 25	En contact
433	En contact téléphonique	En contact
434	Lie	En contact
435	En lien téléphonique	En contact
436	Envoie sms	En contact
437	Contact téléphonique	En contact
438	A comme contact privilégié	En contact
439	Contact téléphonique au sujet code 25 "Jane"	En contact
440	Entre en contact pour signaler sa présence rue jourdain d'Abbans a paris	En contact
441	Est lie	Lien
442	Lie a l'adresse	Lien
443	Lie (2014)	Lien
444	Contacte pour venir récupérer "40" à l'épicerie distri jourdain	En contact
445	Contact pour un "code 42"	En contact
446	Appartient	Titulaire

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
447	Lié à	Lien
448	Lié avec	Titulaire
449	Contacte (probablement Benshaf Kévin)	En contact
450	Contacte	En contact
451	Utilise pour réservation voyages	Utilisateur
452	Associe au boitier imei 355405075649063 découvert chez	Utilisateur
453	En contact téléphonique entre 15h00 et 16h55	En contact
454	En contact téléphonique entre 15h00 et 16h56	En contact
455	En contact téléphonique entre 15h00 et 16h57	En contact
456	En contact téléphonique entre 15h00 et 16h58	En contact
457	En contact téléphonique entre 15h00 et 16h59	En contact
458	En contact téléphonique entre 15h00 et 16h60	En contact
459	En contact téléphonique entre 15h00 et 16h61	En contact
460	En contact téléphonique entre 15h00 et 16h62	En contact
461	Retrouve au domicile	Localisation
462	En contact régulier	En contact
463	Utilise au Maroc	Utilisateur
464	Sms échange avec Lola	En contact
465	A déposé 100 000 euros	Échange objet ou argent
466	Se trouve dans répertoire téléphonique	En contact
467	Se trouve dans répertoire téléphone	En contact
468	Se trouve dans le répertoire téléphone	En contact
469	En communication	En contact
470	A dépose 80 000 euros (entre le 07 et le 09/09/2016)	Échange objet ou argent
471	Non identifié	Autre

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
472	Correspondant privilégié	En contact
473	Client communiqué par tel +213550845563	Utilisateur
474	Utilise par	Utilisateur
475	Nouvelle adresse	Adresse
476	A comme numéro	Utilisateur
477	Même adresse	Même adresse
478	Adresse d'un véhicule utilise	Lien
479	Ancienne adresse (taj)	Adresse
480	Entrée bâtiment b	Adresse
481	Entrée bâtiment a	Adresse
482	Adresse	Adresse
483	Se stationne derrière	Localisation
484	Déclenche un point de position	Localisation
485	Se stationne à côté	Localisation
486	S'engage dans le parking souterrain	Localisation
487	Se trouve dans le box	Localisation
488	Quitte les lieux	Localisation
489	Est stationne devant épicerie	Localisation
490	Utilise	Utilisateur
491	Se trouve avec	Utilisateur
492	Stationne à côté	Localisation
493	Se stationne et téléphone	Localisation
494	Stationne devant	Localisation
495	Se stationne (plusieurs individus à bord)	Localisation
496	Cède	Titulaire

N° Ligne	LIBELLE LIEN	CATEGORIE LIEN
497	Se stationne résidence le Pierront	Localisation
498	Stationne place 638 niveau -2 (appartement 260)	Localisation
499	Se trouve sur le parking au niveau du 70-92	Localisation
500	Se stationne	Localisation
501	Se stationne dans le parking souterrain	Localisation
502	Se stationne à côté	Localisation
503	Stationne dans le parking sous terrain	Localisation
504	Stationne place 639 niveau -2 (appartement 261)	Localisation
505	Est stationne au niveau du n°16	Localisation
506	Entre dans l'impasse	Localisation
507	Se stationne puis une femme monte passager	Localisation
508	Se stationne à la station bp situe à côté	Localisation
509	Entre dans parking souterrain	Localisation
510	Se stationne près	Localisation
511	Se trouve près	Localisation
512	Découvert	Localisation
513	Saisi lors perquisition	Localisation
514	Équipe	Lien
515	Nom inscrit sur bal	Adresse
516	Demeure (agdref, siv)	Adresse
517	Demeure (siv, pc)	Adresse
518	Demeure (engie, fnpc, taj)	Adresse

Annexe 4 Circulaire ministérielle n°30053 du 3 septembre 2019 relative à l'export de données à destination de la recherche scientifique

BULLETIN OFFICIEL DU MINISTÈRE DE L'INTÉRIEUR

MINISTÈRE DE L'INTÉRIEUR

Direction générale
de la gendarmerie nationale

Circulaire n° 30053 du 3 septembre 2019 relative à l'export de données à destination de la recherche scientifique

NOR : INTJ1916834C

Références:

- Code de la sécurité intérieure, notamment ses articles R. 434-8 et R. 434-21;
- Code pénal, notamment son article 226-13;
- Code de procédure pénale, notamment ses articles 11 et R. 156;
- Loi n° 78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés, notamment son article 2;
- Arrêté du 5 septembre 2011 relatif aux règles générales d'emploi des moyens informatiques et des traitements automatisés de données à caractère personnel dans la gendarmerie nationale (NOR: IOJ1129972A);
- Circulaire n° 94855/GEND/CAB du 15 septembre 2011 relative aux règles générales d'emploi des moyens informatiques et des traitements automatisés de données à caractère personnel dans la gendarmerie nationale;
- Charte du gendarme n° 17347 du 15 février 2010.

Pièces jointes : Deux annexes.

La direction générale de la gendarmerie nationale a l'ambition de conforter l'assise scientifique de ses missions. La valorisation de ses liens avec le monde de la recherche publique, para-publique ou privée en constitue un des aspects essentiels.

S'inscrivant ainsi dans la logique d'une force de sécurité intérieure guidée par les éléments probants de la recherche (désignée dans le monde anglo-saxon sous le nom de *evidence-based policing*), la gendarmerie nationale est susceptible de tirer parti de la collecte de ces savoirs pour mieux comprendre les phénomènes criminels, et donc adapter son activité.

La notion de « données » recouvre l'ensemble des données issues des activités de la gendarmerie nationale, notamment opérationnelles, techniques ou de ressources humaines.

Ces données peuvent revêtir une sensibilité particulière et doivent donc être particulièrement protégées avant toute utilisation (consultation ou export) par les chercheurs dans le cadre de leurs travaux.

Le champ d'application de la présente circulaire ne porte ni sur le traitement des données à caractère personnel au sens de la loi n° 78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés et du règlement (UE) n° 2016/679 du Parlement européen et du conseil du 27 avril 2016 relatif à la protection des données (RGPD), ni sur les archives relevant du code du patrimoine.

S'agissant des bases de données statistiques relatives à la criminalité et à la délinquance fournies directement aux chercheurs par le service statistique ministériel de la sécurité intérieure (SSMSI), les procédures propres à ce service s'appliqueront, dans le respect des règles de confidentialité et de sécurité des données, et en coordination avec la direction générale de la gendarmerie nationale.

La présente circulaire encadre le recours à des partenariats de recherche par convention pour l'exploitation de données entièrement anonymisées et protégées (ne revêtant pas de caractère personnel) et prévoit une habilitation individuelle de chaque chercheur par un accord de confidentialité.

1. La mise en place de partenariats de recherche

1.1. L'encadrement institutionnel: la convention de partenariat

La mise à disposition de données de la gendarmerie et leur exploitation par les chercheurs ont pour préalable la conclusion d'une convention avec les organismes partenaires dont ces derniers dépendent.

Fondée sur le modèle proposé en annexe I, cette convention définit clairement les objectifs et les bénéfices mutuels attendus du partenariat. Elle en décrit les modalités, les échéances, la responsabilité des parties, les règles de confidentialité, de publication et de propriété intellectuelle.

15 SEPTEMBRE 2019. – INTÉRIEUR 2019-9 – PAGE 1

Cette convention apporte les garanties d'un transfert encadré de données et porte une attention particulière à la préservation du secret professionnel, d'une part, et à la suppression du caractère identifiant des données, d'autre part.

Dans l'hypothèse où les données ont un caractère personnel, la convention prévoit notamment la mise en œuvre d'un processus d'anonymisation afin que celles confiées au chercheur soient dépourvues de tout élément permettant d'identifier directement ou indirectement des personnes physiques. L'objectif est d'assurer la légalité du recueil, du traitement et de l'exploitation des données, ainsi que la fiabilité des conclusions qui en émanent.

1.2. *L'encadrement personnel: l'accord de confidentialité*

Au terme d'une première phase d'évaluation du contexte de l'étude (le cas échéant avec élaboration d'un protocole de recherche), la convention de partenariat entre la gendarmerie nationale et le partenaire est ensuite complétée par un accord de confidentialité entre le chercheur et son service d'accueil de la gendarmerie.

Plus précisément, cet accord de confidentialité indique la nature des données remises, les techniques d'anonymisation utilisées et les modalités d'exploitation. Un exemple de ce type d'accord est en annexe II de la présente circulaire.

Cet accord prévoit le suivi du projet de recherche par un officier référent individuellement désigné par le commandant d'une formation administrative ou assimilé. Spécialement sensibilisé à cette mission, cet officier est chargé de conseiller et d'accompagner le travail du chercheur et de veiller au respect du secret professionnel et de la protection des données à caractère personnel.

Cet officier a pour mission :

- d'évaluer le besoin, notamment en termes d'anonymisation et de sécurisation;
- de procéder ou faire procéder à la collecte des données;
- de garantir leur anonymat;
- de procéder au transfert sécurisé de la donnée aux chercheurs;
- de s'assurer du respect par les chercheurs des conditions de sécurisation;
- d'accompagner le traitement statistique des données et leur interprétation;
- de prendre connaissance des publications;

afin de garantir la légalité des analyses et des éléments restitués.

La convention de partenariat est transmise au secrétariat général de l'Observatoire national des sciences et technologies de la sécurité (ONSTS). Elle est examinée par un comité consultatif de valorisation de la donnée qui exprime un avis sur la conformité et l'opportunité du projet.

Ce comité se compose de représentants :

- de l'Observatoire national des sciences et technologies de la sécurité;
- du cabinet du directeur général de la gendarmerie nationale;
- de la mission du pilotage et de la performance;
- de la direction des opérations et de l'emploi;
- du service des technologies et des systèmes d'information de la sécurité intérieure;
- de la direction des soutiens et des finances;
- de la direction des personnels militaires de la gendarmerie nationale;
- du pôle judiciaire de la gendarmerie nationale;
- du centre de recherche de l'école des officiers de la gendarmerie nationale;
- du service statistique ministériel de la sécurité intérieure, pour l'étude de la mise à disposition des données statistiques relatives à la criminalité et à la délinquance.

La convention, accompagnée de l'avis du comité, est ensuite soumise à la validation de l'administrateur des données de la gendarmerie (mission du pilotage et de la performance) pour signature du directeur général de la gendarmerie nationale.

L'accord de confidentialité est signé par le chef du service d'accueil, l'officier référent, ainsi que chacun des chercheurs appelés à travailler sur le projet.

2. **L'exploitation des données**

2.1. *La collecte des données par la gendarmerie nationale*

Le partenariat de recherche repose sur la collecte de données de la gendarmerie nationale. Ces données sont issues :

- des observations et des entretiens menés au sein de ses unités ou de ses services;
- des documents et des procédures détenus par ses unités ou ses services;
- des traitements de données à caractère personnel mis en œuvre par ses unités ou ses services.

Les observations réalisées sur site doivent respecter les conditions précisées par le commandant d'unité ou le chef de service concerné. Elles doivent notamment respecter les principes de discrétion et de secret professionnel.

Les entretiens avec les militaires de la gendarmerie doivent également s'inscrire dans ces principes, recueillir le consentement du militaire et respecter les limites de confidentialité imposées par celui-ci.

Les documents et procédures peuvent être exploités par le chercheur sous condition d'acceptation du responsable de la donnée. S'agissant des procédures judiciaires, une demande d'autorisation devra être réalisée auprès du procureur de la République territorialement compétent. Cette demande est motivée et précise notamment l'objet de la recherche, sa durée, l'identité du chercheur, ainsi que tout élément justifiant le besoin d'exploitation de ces procédures. La demande indique également que la gendarmerie s'engage à exporter au chercheur uniquement des pièces de procédure judiciaire anonymisées (deux hypothèses sont proposées au parquet : soit une anonymisation par les services de gendarmerie à l'origine de la demande avant transmission au chercheur ; soit une anonymisation par le parquet).

La collecte de données issues des traitements de données à caractère personnel ne peut être réalisée que par un accédant habilité, dans les conditions prévues par le texte réglementaire relatif à ce traitement. En tant que de besoin, il peut être soutenu par l'exploitant de la base.

En toute hypothèse, sauf si cela est expressément motivé au regard de l'objet de la recherche, la collecte ne peut concerner des éventuelles données relatives à la prétendue origine raciale ou l'origine ethnique, aux opinions politiques, aux convictions religieuses ou philosophiques, à l'appartenance syndicale, à la biométrie, à la santé, à la vie sexuelle ou à l'orientation sexuelle d'une personne physique.

2.2. L'anonymisation des données par la gendarmerie nationale

Les données intéressant la recherche relevant du régime de protection des données à caractère personnel doivent faire l'objet d'une précaution spécifique. Sur la base des éléments collectés, il s'agit de supprimer les éléments « identifiants » par un procédé d'anonymisation.

Cette opération est réalisée par un personnel de la gendarmerie nationale (spécialement habilité s'il s'agit d'un traitement de données à caractère personnel) sous la responsabilité de l'officier référent. Elle consiste à supprimer, remplacer ou appauvrir les données de sorte que, de façon unitaire ou collective, elles ne puissent constituer directement ou indirectement des données à caractère personnel.

Ces techniques d'anonymisation sont mises en œuvre dans le respect des normes AFNOR afférentes et prévues sous les références ISO 20889 et ISO 29100 en vigueur.

Exemple : processus d'anonymisation de données extraites de comptes-rendus de police judiciaire (CRPJ).

DONNÉES SOURCES	ACTION	DONNÉES ANONYMISÉES
UNA	Remplacement	Référentiel propre à la base de donnée
Service saisi	Appauvrissement	Catégorie de service saisi (CIAT, BT, SUD, BR, DIPJ, SR, office...), le cas échéant complété par l'aire géographique sur laquelle se situe ce service
Date des faits	Appauvrissement	En fonction de la rareté du contentieux et des nécessités pour l'analyse, suppression du jour, voire du mois des faits
Manière d'opérer	Appauvrissement	Purge de l'ensemble des noms, prénoms, adresses et autres identifiants subsistant dans les libellés
Lieu des faits (rue, commune, dépt)	Appauvrissement	Catégorie de commune (par ex. en fonction de la population)+ département
Nom, prénom victime et mis en cause	Suppression	-
Nom, prénom enquêteur	Appauvrissement	Catégorie d'enquêteur (par ex. grade, qualité, expérience, charge de travail, unité d'affectation...)
Adresse (rue, commune, dépt) de la victime ou de l'auteur	Appauvrissement	Département
Genre victime et mis en cause	Maintien	Genre
Date de naissance victime et mis en cause	Appauvrissement	Age
Nationalité victime et mis en cause	Maintien	Nationalité
Profession victime et mis en cause	Appauvrissement	Catégorie socio-professionnelle
Antécédents judiciaires mise en cause	Appauvrissement	Catégorie d'infraction

Ainsi anonymisées, les données sont remises par l'officier référent au chercheur. À ce stade, aucune donnée à caractère personnel ne doit subsister dans les données remises.

Le chercheur pourra alors procéder au traitement des données anonymisées remises selon les méthodes propres à sa matière et comprenant généralement les étapes de :

- nettoyage: correction ou suppression des valeurs aberrantes ou inexactes;
- uniformisation: transformation des données afin de les rendre comparables;
- structuration: organisation des données selon une configuration utile au projet de recherche;
- qualification: ajout d'informations utiles à la compréhension de la donnée (par ex. confidentialité, fiabilité...).

L'export des données par l'officier référent est réalisé dans un cadre sécurisé.

Les extractions de données non anonymisées sont conservées par l'officier référent durant le déroulement du projet de recherche, ou un personnel de la gendarmerie nationale agissant sous sa responsabilité. L'officier référent et ce personnel doivent être nommément désignés et spécialement habilités à accéder au traitement dont les données sont issues. Il est possible d'y avoir recours afin de procéder aux levées de doutes et corrections nécessaires lors des phases de nettoyage des données et d'analyse. Cette conservation est établie de manière sécurisée et, à l'issue du projet de recherche, ces données sont définitivement supprimées.

Le correspondant de la gendarmerie nationale auprès du délégué à la protection des données du ministère de l'intérieur dispose d'un droit d'accès et de contrôle sur les procédés employés aux fins d'exclure des données tout élément à caractère personnel, en particulier les procédés d'anonymisation.

2.3. La transmission des données au chercheur

La transmission des données par l'officier référent est réalisée dans un cadre sécurisé tel que défini dans la convention. Sont uniquement transmises des données anonymisées.

S'agissant des actes de procédures judiciaires, seules les informations dont la communication est autorisée par le procureur de la République sont transmises au chercheur dans le cadre de ses travaux.

2.4. La détention et l'analyse des données par le chercheur

Une fois en possession des données, le chercheur est tenu de prendre les précautions requises par la convention quant à leur détention et à leur usage. Lorsque les données sont considérées comme sensibles et lorsque le chercheur est amené à les exploiter sur son propre système, la convention pourra prévoir la mise en œuvre d'un dispositif de protection par chiffrement selon des modalités conformes à l'état de l'art. En aucun cas, ces données ne peuvent être cédées à, ou exploitées par, des personnes extérieures à la convention de partenariat.

Réalisée à des fins statistiques, scientifiques ou historiques, l'analyse des données peut recourir à des méthodes quantitatives ou qualitatives.

Celles-ci relèvent de la compétence du chercheur, qui dispose des savoir-faire et des outils nécessaires à l'exploitation des données. L'analyse est néanmoins réalisée en lien avec l'officier référent au titre de son rôle de conseil et d'accompagnement.

Ce travail collaboratif entre le chercheur et le professionnel permet une analyse pertinente des données, notamment :

- une pondération de la fiabilité des informations en fonction des processus métiers ayant abouti à leur collecte, des dispositifs techniques de remontée de l'information, des modes de requête de la base et des biais d'analyse;
- un nouvel examen du respect des règles de secret.

Sans préjudice de la liberté d'expression et des droits de propriété intellectuelle, le chercheur communique le résultat de ses travaux à l'officier référent. Si le chercheur en fait la demande, l'officier référent le conseille sur l'intérêt et l'opportunité de ses analyses.

Cette communication s'applique également aux publications ultérieures utilisant les données collectées.

3. La conservation des données

3.1. Le stockage des données

Selon la nature des recherches envisagées, la convention pourra prévoir une rétrocession par le chercheur des différentes versions de son travail. Ces versions reflètent les différentes étapes de traitement de la donnée :

- base de données brutes et anonymisées;
- base de données nettoyées;
- traitement qualitatif (c'est-à-dire encodage des données);
- traitement quantitatif (analyse univariée);
- traitement quantitatif (analyse bivariée);
- traitement quantitatif (analyse multivariée, avec une version pour chaque modèle probabiliste mis en œuvre);
- traitement en intelligence artificielle (version des résultats en fonction des modèles appliqués);
- analyses (en leurs différentes étapes);
- et résultats (en ses différentes dimensions).

Cette préservation des versions est réalisée à des fins de traçabilité, de contrôle ou de réemploi. Elle ne constitue pas une contrainte particulière dans la mesure où elle relève de la démarche méthodologique habituellement employée par les chercheurs.

Les versions ainsi collectées sont organisées de façon claire et accessible sur un support sécurisé dont les accès sont restreints par un dispositif de gestion des droits d'accès. Le stockage technique des données est hébergé par le STIG.

3.2. *L'administration des données*

L'administration des données ainsi conservées est confiée au responsable de la recherche et du développement au sein du service d'accueil, en lien avec l'administrateur des données gendarmerie.

Ces données peuvent être réutilisées en tant que :

- moyen de contrôle, voire de contre-expertise d'une étude ;
- ressource pour une nouvelle étude.

La présente circulaire sera publiée au *Bulletin officiel* du ministère de l'intérieur.

Fait le 3 septembre 2019.

Le général d'armée,
directeur général de la gendarmerie nationale,
R. LIZUREY

Annexe 5 Liste, occurrences et fréquence des symboles terminaux existant dans les données prétraitées

Symbole terminal	Occurrences	Fréquence
PERSONNE-adresse-LIEU (ADRESSE)	293	12,43%
PERSONNE-utilisateur-VÉHICULE	261	11,07%
PERSONNE-titulaire-VÉHICULE	177	7,51%
VÉHICULE-localisation-LIEU (ADRESSE)	169	7,17%
PERSONNE-familial amical-PERSONNE	162	6,87%
PERSONNE-en contact-PERSONNE	155	6,58%
PERSONNE-utilisateur-TÉLÉCOM	137	5,81%
PERSONNE MORALE-adresse-LIEU (ADRESSE)	97	4,12%
PERSONNE-échange objet ou argent-PERSONNE	92	3,90%
PERSONNE-titulaire-TÉLÉCOM	92	3,90%
PERSONNE-professionnel-PERSONNE MORALE	92	3,90%
PERSONNE-localisation-LIEU PUBLIC	81	3,44%
TÉLÉCOM-en contact-TÉLÉCOM	65	2,76%
PERSONNE-localisation-LIEU (ADRESSE)	55	2,33%
VÉHICULE-localisation-LIEU PUBLIC	48	2,04%
PERSONNE-criminel-PERSONNE	35	1,48%
PERSONNE MORALE-titulaire-VÉHICULE	33	1,40%
PERSONNE-est-PERSONNE	31	1,32%
TÉLÉCOM-en contact-PERSONNE	25	1,06%
LIEU PUBLIC-adresse-LIEU (ADRESSE)	24	1,02%
PERSONNE-titulaire-LIEU (ADRESSE)	18	0,76%
PERSONNE-échange objet ou argent-VÉHICULE	18	0,76%
PERSONNE-localisation-VÉHICULE	16	0,68%
VÉHICULE-localisation-VÉHICULE	15	0,64%
VÉHICULE-localisation-LIEU PRIVÉ	12	0,51%
PERSONNE-professionnel-LIEU PUBLIC	12	0,51%
PERSONNE MORALE-titulaire-TÉLÉCOM	11	0,47%
PERSONNE MORALE-titulaire-LIEU (ADRESSE)	11	0,47%
TÉLÉCOM-lien-LIEU (ADRESSE)	11	0,47%
PERSONNE MORALE-utilisateur-VÉHICULE	7	0,30%
PERSONNE-financier-PERSONNE MORALE	7	0,30%
PERSONNE-fait partie-PERSONNE	6	0,25%
PERSONNE-professionnel-PERSONNE	5	0,21%
TÉLÉCOM-localisation-LIEU (ADRESSE)	5	0,21%
PERSONNE-M.E.C.-OBJET	5	0,21%
PERSONNE-adresse-PERSONNE	4	0,17%
PERSONNE-même nom-PERSONNE	4	0,17%
PERSONNE-adresse-LIEU PRIVÉ	4	0,17%
PERSONNE-localisation-PERSONNE MORALE	4	0,17%
PERSONNE-échange objet ou argent-LIEU (ADRESSE)	4	0,17%

Symbole terminal	Occurrences	Fréquence
PERSONNE-échange objet ou argent-LIEU PUBLIC	4	0,17%
OBJET-lien-VÉHICULE	4	0,17%
PERSONNE-utilisateur-PERSONNE	3	0,13%
LIEU PUBLIC-titulaire-VÉHICULE	3	0,13%
PERSONNE-professionnel-LIEU (ADRESSE)	3	0,13%
PERSONNE-autre-LIEU (ADRESSE)	3	0,13%
PERSONNE-autre-VÉHICULE	3	0,13%
LIEU PRIVÉ-adresse-LIEU (ADRESSE)	3	0,13%
PERSONNE-lien interdit-PERSONNE	2	0,08%
PERSONNE-adresse-PERSONNE MORALE	2	0,08%
VÉHICULE-localisation-PERSONNE MORALE	2	0,08%
PERSONNE-titulaire-PERSONNE MORALE	2	0,08%
TÉLÉCOM-lien-PERSONNE	2	0,08%
VÉHICULE-lien-LIEU (ADRESSE)	2	0,08%
PERSONNE-utilisateur-BANQUE	2	0,08%
PERSONNE-non lien-PERSONNE	1	0,04%
PERSONNE-adresse-LIEU PUBLIC	1	0,04%
PERSONNE-utilisateur-LIEU PRIVÉ	1	0,04%
PERSONNE-titulaire-OBJET	1	0,04%
LIEU PUBLIC-titulaire-TÉLÉCOM	1	0,04%
PERSONNE-localisation-LIEU PRIVÉ	1	0,04%
PERSONNE-échange objet ou argent-LIEU PRIVÉ	1	0,04%
OBJET-localisation-VÉHICULE	1	0,04%
TÉLÉCOM-localisation-VÉHICULE	1	0,04%
OBJET-localisation-LIEU (ADRESSE)	1	0,04%
PERSONNE MORALE-professionnel-LIEU (ADRESSE)	1	0,04%
PERSONNE-antécédents-OBJET	1	0,04%
PERSONNE MORALE-autre-LIEU (ADRESSE)	1	0,04%
TÉLÉCOM-autre-TÉLÉCOM	1	0,04%
TOTAL	2 357	100%

Légende : Sont surlignés les symboles terminaux soumis au questionnaire

Annexe 6 Questionnaire soumis aux analystes de la Gendarmerie nationale pour appuyer l'attribution des forces liantes des symboles terminaux

Pondération de liens - Maud BENICHO DUHIL de BENAZE

Annexe : Questionnaire

Pondération de liens - Maud BENICHO DUHIL de BENAZE

Madame, Monsieur

Veuillez prendre quelques minutes de votre temps

pour remplir le sondage suivant.

Ce questionnaire s'inscrit dans le cadre d'une étude en partenariat avec le Service Central de Renseignement Criminel de la Gendarmerie Nationale (SCRC). Elle a pour objet de modéliser les groupes criminels organisés (GCO).

A cet effet, elle cherche à comprendre les relations existantes entre les membres d'une organisation criminelle à partir des données d'enquête (liens entre différentes entités, telles que des téléphones, véhicules, objets, lieux, personnes, etc). Le but est de permettre le développement d'une méthode visant à identifier les individus ayant rôles et positions clés dans un réseau.

Il s'agit de prendre en compte des critères qui n'existent pas actuellement dans les schémas anacrim tels que la pondération des liens sociaux entre les individus (importances des acteurs dans le réseau, compétences, jeux de pouvoir, etc).

Dans mon étude, je cherche notamment à déterminer si deux individus reliés dans le graphe par une chaîne de liens et d'entités sont effectivement en contact.

Par exemple, une chaîne de type :

Chaîne 1 : **MATHILDE-possède-0612345678-appelle-0698765432-appartient à-FRANCK** permet de lier directement Mathilde et Franck.

En revanche, une chaîne de type :

Chaîne 2 : **MATHILDE-adresse-3 RUE DE L'ARBRE-siège social de-SOCIETE MACHIN-emploie-Franck** permet pas de lier directement Mathilde et Franck avec fiabilité.

La méthode employée consiste à découper les chaînes de liens en symboles (**TYPE ENTITE - type lien - TYPE ENTITE**) et attribuer à chaque symbole une pondération destinée à indiquer l'intensité du lien entre deux entités.

Les différents symboles de chaque chaîne de liens seront ensuite combinés pour obtenir le niveau de confiance dans le lien direct obtenu entre deux individus, quel que soit le nombre d'intermédiaires les séparant.

Pour reprendre les exemples précédents, on aurait (les valeurs sont attribuées à titre d'exemple, la combinaison des liens intermédiaires établit le niveau de confiance entre les deux individus) :

1. Poids(PERSONNE-propriété-TELECOM) = 5; Poids(TELECOM-en contact-TELECOM)= 8

Poids(Chaîne 1) = Poids(PERSONNE-propriété-TELECOM) + Poids(TELECOM-en contact-TELECOM) + Poids(PERSONNE

propriété-TELECOM) = (5+8+5) = 18 - FORT NIVEAU DE CONFIANCE dans l'existence du lien Mathilde - Franck

2. Poids(PERSONNE-adresse-LIEU) = 4; POIDS(LIEU-adresse-PERSONNE MORALE)=1; Poids(PERSONNE MORALE-professionnel-PERSONNE) = 2

Poids(Chaîne 2) = Poids(PERSONNE-adresse-LIEU) + POIDS(LIEU-adresse-PERSONNE MORALE) + Poids(PERSONNE MORALE-professionnel-PERSONNE) = (4+1+2) = 7 - FAIBLE NIVEAU DE CONFIANCE dans l'existence du lien Mathilde - Franck

Ce questionnaire s'adresse aux enquêteurs et analystes de la gendarmerie. Il s'appuie sur leur expérience pour définir les pondérations à attribuer aux différents liens intermédiaires.

Cette étude s'inscrit dans le contexte d'un GCO trafiquant de l'argent liquide (collectant l'argent de ses "clients" afin de le rapatrier vers leurs fournisseurs notamment).

Afin que nous parlions tous de la même chose, je me permets ici une rapide distinction entre certains libellés de liens que vous trouverez dans les 7 questions :

* *propriétaire* : la seule notion de propriété est prise en compte (appartement, ligne téléphonique)

* *utilisateur* : à la différence de la propriété, la seule notion d'utilisation est à prendre en compte (ex : utiliser un véhicule qui appartient à quelqu'un d'autre)

* *adresse* : il n'y a pas forcément de propriété du logement dans lequel vit un individu

* *localisation* : individu ou objet dont l'enquête a déterminé la présence à un endroit à un moment donné (différent de l'adresse)

* *professionnel* : le lien établi par l'enquête entre les deux objets/personnes concernées est de nature professionnelle

* *échange objet ou argent* : dépôt ou retrait d'un sac fermé, ou d'argent liquide

* *en contact* : il peut s'agir d'échanges téléphoniques, verbaux, etc, la notion est assez large

Vos réponses sont très précieuses pour la validité de cette étude.

Mille mercis pour votre participation !

1 Dans le cadre d'une enquête sur un GCO, quel poids attribueriez-vous aux types de liens suivants ?

Propositions de réponse: *Choisissez une seule réponse dans chaque rangée*

	0 - lien négligeable	2 - lien faible	4 - lien potentiellement fort	6 - lien fort	8 - lien très fort	10 - lien extrêmement fort
PERSONNE -> adresse -> LIEU (caractère public/privé non précisé)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PERSONNE -> localisation -> LIEU PRIVE	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PERSONNE -> localisation -> LIEU PUBLIC	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PERSONNE -> localisation -> LIEU (caractère public/privé non précisé)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PERSONNE -> propriétaire -> LIEU (caractère public/privé non précisé)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PERSONNE MORALE -> adresse -> LIEU (caractère public/privé non précisé)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

2 Dans le cadre d'une enquête sur un GCO, quel poids attribueriez-vous aux types de liens suivants ?

Propositions de réponse: *Choisissez une seule réponse dans chaque rangée*

	0 - lien négligeable	2 - lien faible	4 - lien potentiellement fort	6 - lien fort	8 - lien très fort	10 - lien extrêmement fort
PERSONNE -> utilisateur -> VEHICULE	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PERSONNE -> propriétaire -> VEHICULE	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PERSONNE -> utilisateur -> TELECOM	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PERSONNE MORALE -> propriétaire -> VEHICULE	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

3 Dans le cadre d'une enquête sur un GCO, quel poids attribueriez-vous aux types de liens suivants ?

Propositions de réponse: *Choisissez une seule réponse dans chaque rangée*

	0 - lien négligeable	2 - lien faible	4 - lien potentiellement fort	6 - lien fort	8 - lien très fort	10 - lien extrêmement fort
VEHICULE -> localisation -> LIEU (caractère public/privé non précisé)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
VEHICULE -> localisation -> LIEU PUBLIC	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
VEHICULE -> localisation -> VEHICULE (deux VHL dont les conducteurs discutent ou échangent)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

4 Dans le cadre d'une enquête sur un GCO, quel poids attribueriez-vous aux types de liens suivants ?

Propositions de réponse: *Choisissez une seule réponse dans chaque rangée*

	0 - lien négligeable	2 - lien faible	4 - lien potentiellement fort	6 - lien fort	8 - lien très fort	10 - lien extrêmement fort
PERSONNE -> échange objet ou argent -> VEHICULE	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PERSONNE -> échange objet ou argent -> PERSONNE	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PERSONNE -> échange objet ou argent -> LIEU PUBLIC	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PERSONNE -> échange objet ou argent -> LIEU (caractère public/privé non précisé)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

5 Dans le cadre d'une enquête sur un GCO, classez les liens suivants entre PERSONNE par ordre de pertinence

Propositions de réponse: *Changez l'ordre selon vos préférences (1. - le plus important, dernier - le moins important)*

PERSONNE-familial/amical-PERSONNE	<input type="text"/>
PERSONNE-en contact-PERSONNE	<input type="text"/>
PERSONNE-antécédents criminels avec-PERSONNE	<input type="text"/>
PERSONNE-est-PERSONNE (deuxième identité)	<input type="text"/>
PERSONNE-fait partie-CLAN	<input type="text"/>
PERSONNE-professionnel-PERSONNE	<input type="text"/>
PERSONNE-vit chez-PERSONNE	<input type="text"/>
PERSONNE-utilisateur-PERSONNE (fausse identité)	<input type="text"/>
PERSONNE-échange objet ou argent-PERSONNE	<input type="text"/>

6 Quelle force attribueriez-vous à un lien "TELECOM -> en contact -> TELECOM" ?

Propositions de réponse: *Choisissez une seule réponse*

- 0 - lien négligeable 1 2 - lien faible 3 4 - lien potentiellement fort 5 6 - lien fort
- 7 8 - lien très fort 9 10 - lien extrêmement fort

7 Quelle force attribueriez-vous à un lien "PERSONNE -> professionnel -> PERSONNE MORALE" ?

Propositions de réponse: *Choisissez une seule réponse*

- 0 - lien négligeable 1 2 - lien faible 3 4 - lien potentiellement fort 5 6 - lien fort
- 7 8 - lien très fort 9 10 - lien extrêmement fort

8 Avez-vous une remarque ? Un conseil ? Une critique ? Ou même plusieurs ?

Merci infiniment pour votre participation !!

Statistique de visites

302

Nombre de visites

128

Questionnaires complets

0

Questionnaires incomplets

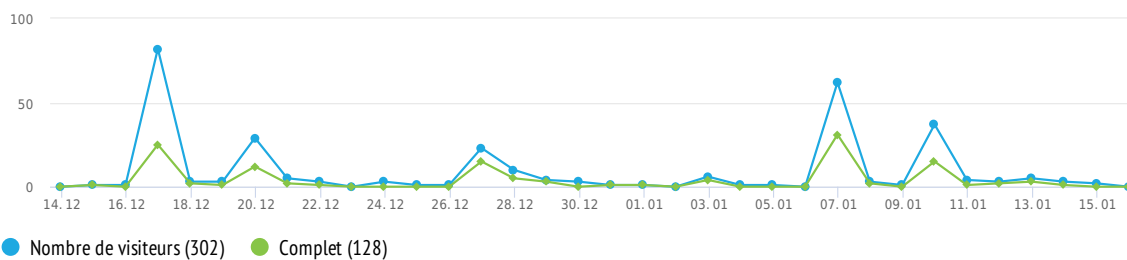
174

Seulement affichés

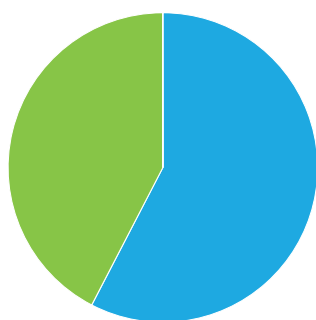
42,4%

Taux de remplissage

Historique des visites (15-12-2021 – 14-01-2022)

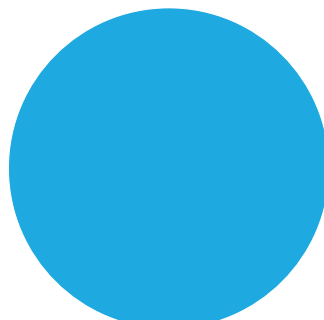


Nombre total de visites



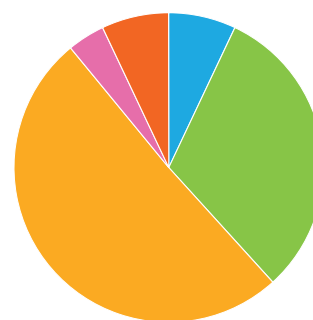
- Seulement affiché (57,6 %)
- Complet (42,4 %)
- Incomplet (0,0 %)

Sources des visites



- Lien direct (100,0 %)

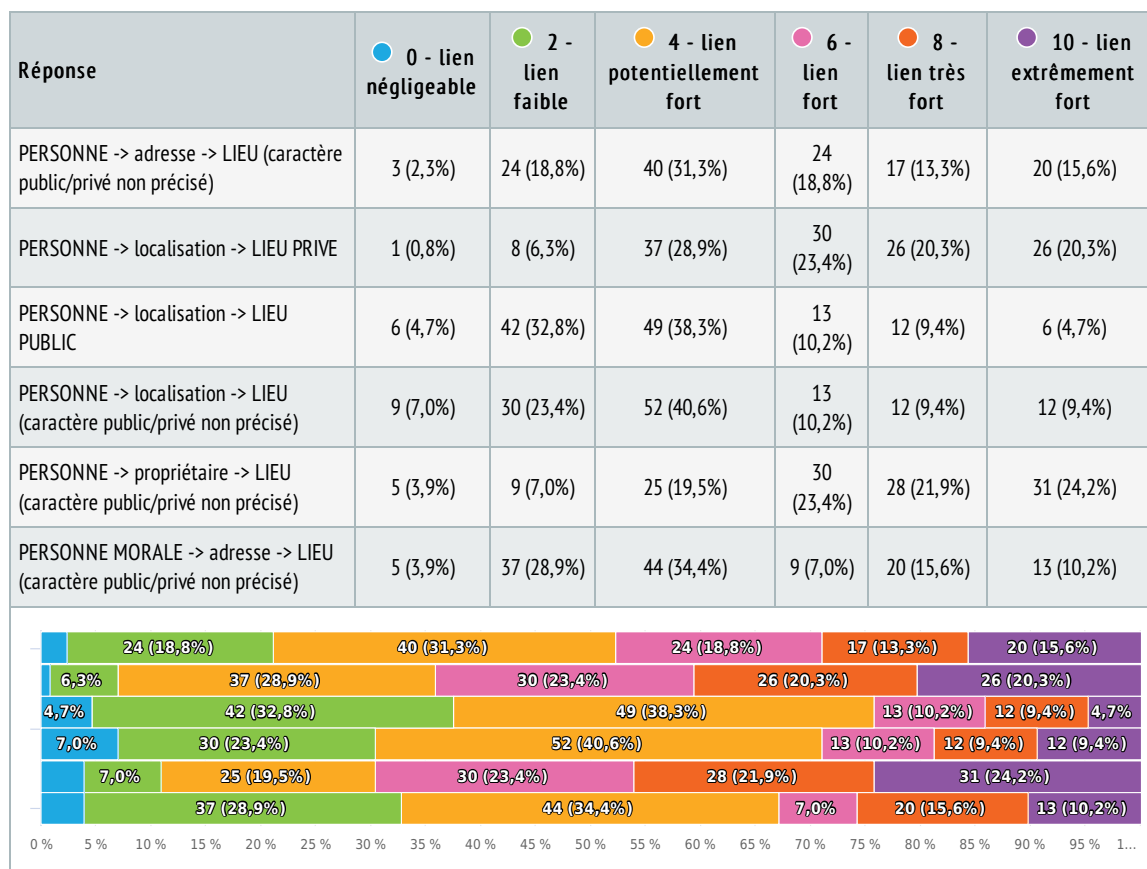
Durée de réalisation



- 2-5 min. (7,0 %)
- 5-10 min. (31,3 %)
- 10-30 min. (50,8 %)
- 30-60 min. (3,9 %)
- >60 min. (7,0 %)

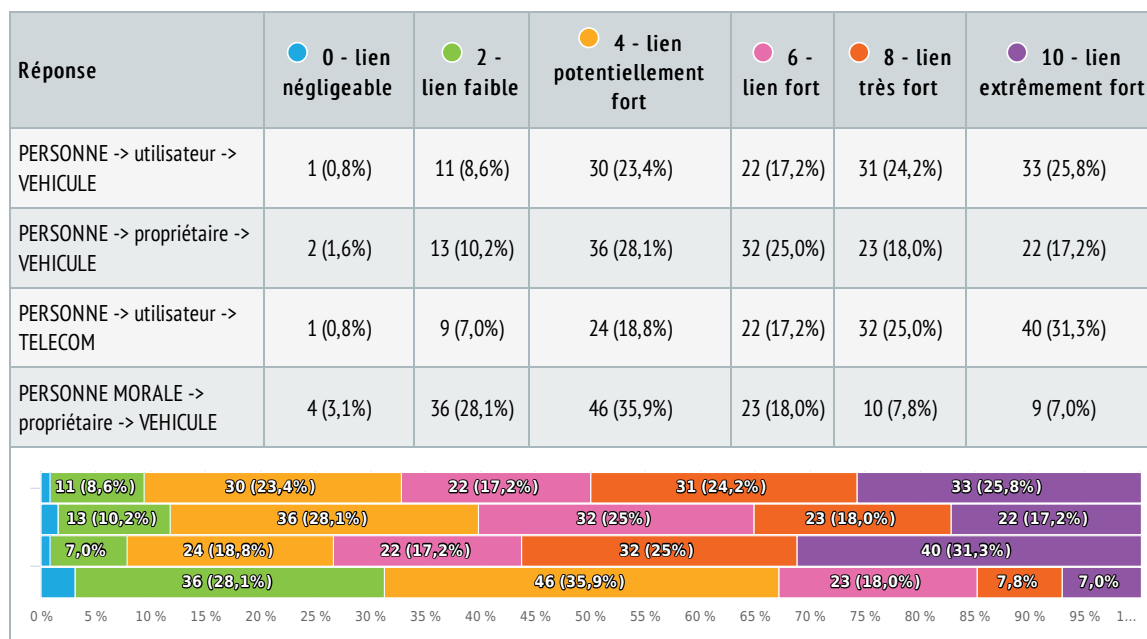
1 Dans le cadre d'une enquête sur un GCO, quel poids attribueriez-vous aux types de liens suivants ?

Matrice de choix uniques, Nombre de répondants 128 x, sans réponse 0 x



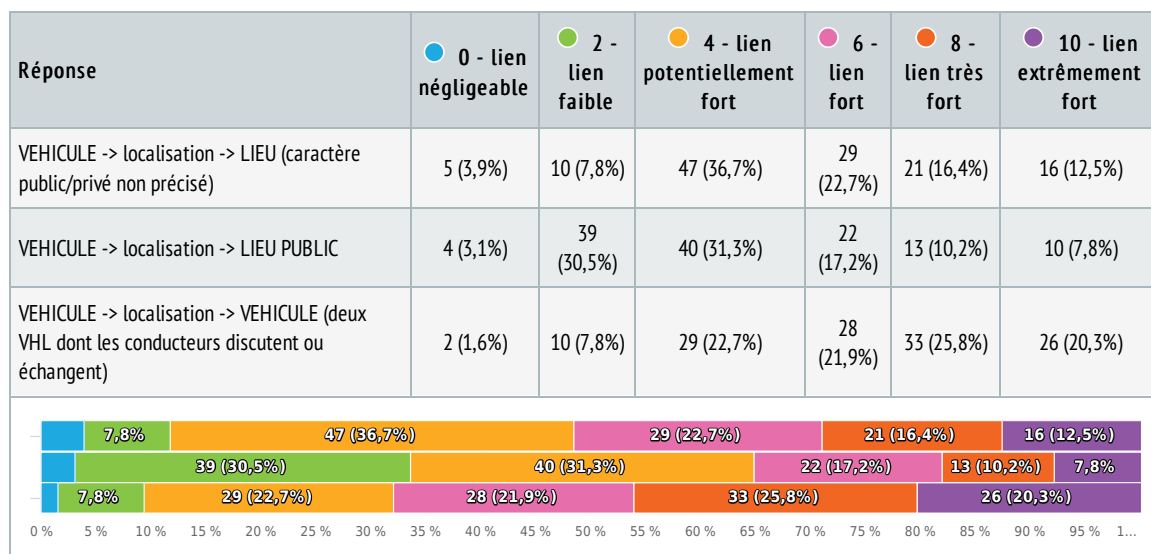
2 Dans le cadre d'une enquête sur un GCO, quel poids attribueriez-vous aux types de liens suivants ?

Matrice de choix uniques, Nombre de répondants 128 x, sans réponse 0 x



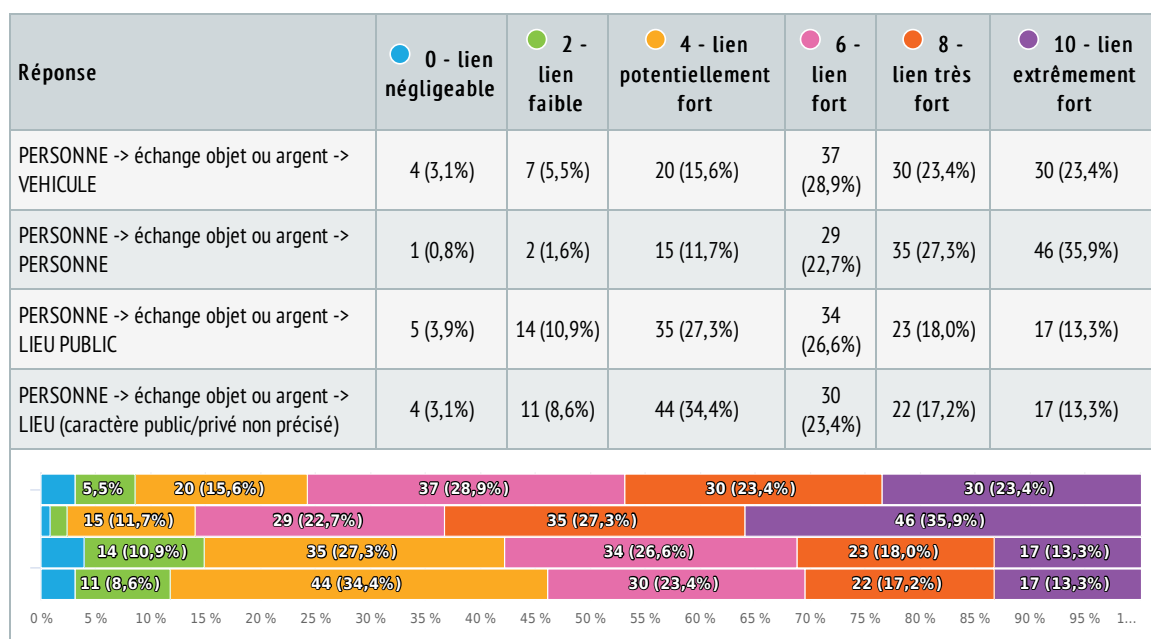
3 Dans le cadre d'une enquête sur un GCO, quel poids attribueriez-vous aux types de liens suivants ?

Matrice de choix uniques, Nombre de répondants 128 x, sans réponse 0 x



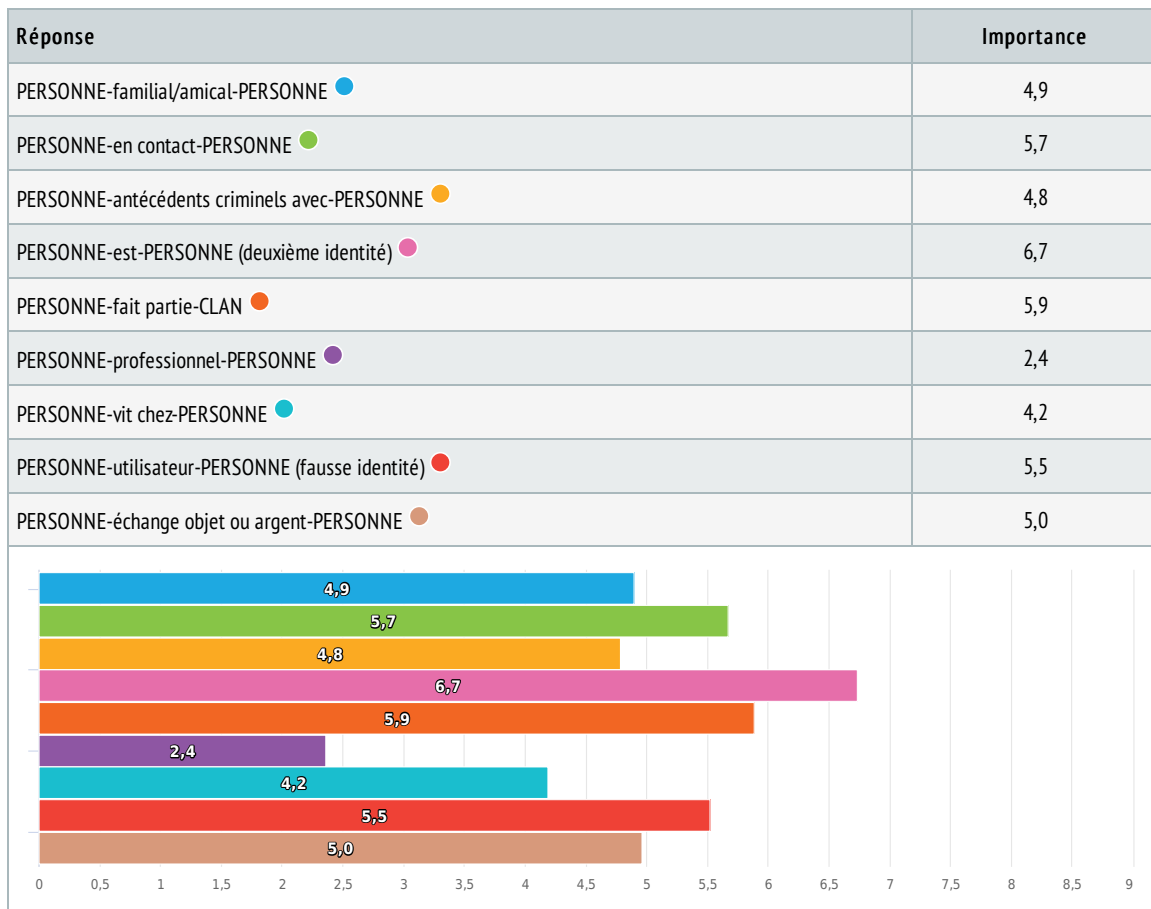
4 Dans le cadre d'une enquête sur un GCO, quel poids attribueriez-vous aux types de liens suivants ?

Matrice de choix uniques, Nombre de répondants 128 x, sans réponse 0 x



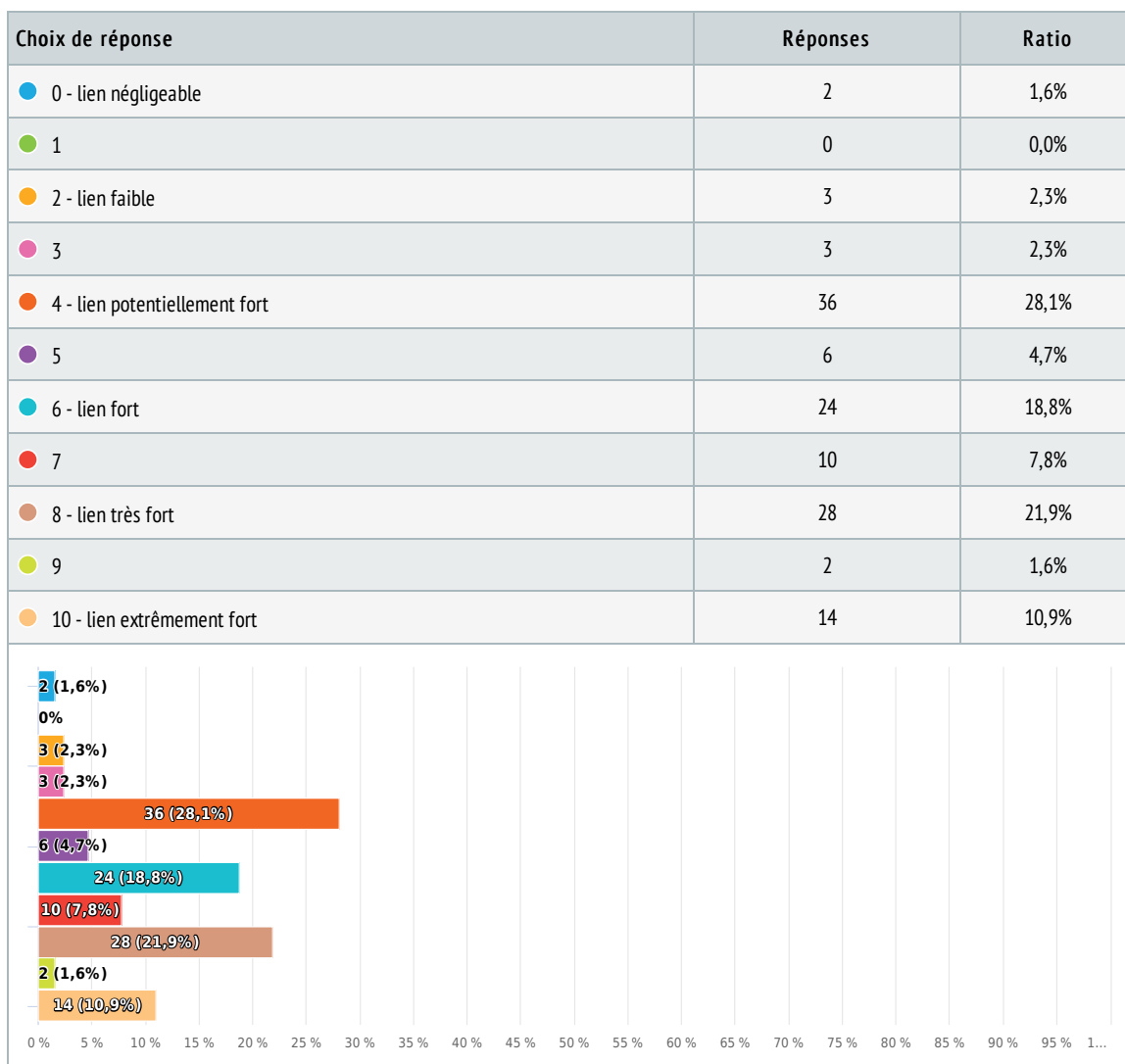
5 Dans le cadre d'une enquête sur un GCO, classez les liens suivants entre PERSONNE par ordre de pertinence

Classement par ordre de préférence, Nombre de répondants 128 x, sans réponse 0 x



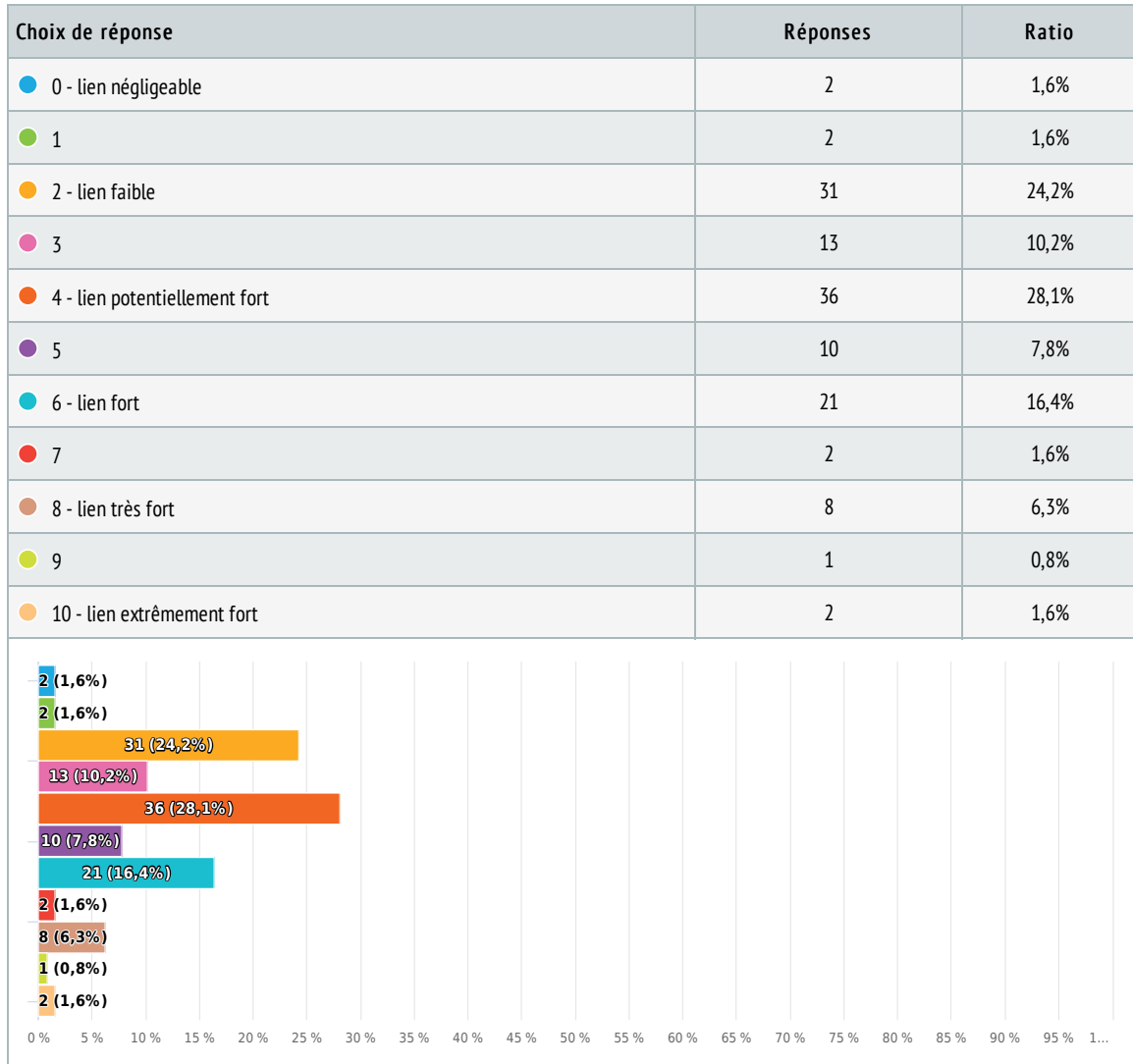
6 Quelle force attribueriez-vous à un lien "TELECOM -> en contact -> TELECOM" ?

Choix unique , Nombre de répondants 128 x, sans réponse 0 x



7 Quelle force attribueriez-vous à un lien "PERSONNE -> professionnel -> PERSONNE MORALE" ?

Choix unique, Nombre de répondants 128 x, sans réponse 0 x



8 Avez-vous une remarque ? Un conseil ? Une critique ? Ou même plusieurs ?

Réponses textuelles , Nombre de répondants 128 x, sans réponse 0 x

- (8x)
-
- (4x).
- -
- (2x)/
- Attention certains chefs de GCO meetent otu en place pour ne pas avoir de lien avec le groupe. Certains signaux faibles sont à prendre en considération
- Aucune
- aucune remarque
- (2x) Aucune remarque
- Aucune remarque.
- bon courage ;)
- Bon courage à vous
- Bonjour, bon courage !
- Bonjour, exercice très difficile sortie du contexte. Apprécier un "poids" pour des liens génériques est à mon sens source d'erreur. Cette donnée doit être validée par les enquêteurs et/ou Anacrim au moment des constatations car d'un cas à l'autre il peut y avoir des différences. Par exemple pour certains la téléphonie sera une valeur sure et pour d'autre elle n'aura aucune valeur car peu attaché à ce mode de communication. Il en va de même pour un véhicule et une adresse.
- Bonjour, j'ai pondéré mes réponses à minima. Tous les liens que nous pouvons faire lors d'enquête sont par nature extrêmement faibles et sont sujets à interprétation. L'emploi du conditionnel est donc à privilégier dans les actes d'analyse que nous faisons au quotidien.
- Bonjour, je me permets seulement de vous faire part que je ne suis pas tout à fait d'accord avec vos conclusions pour la chaîne 1. Les titulaires des lignes ne sont pas toujours et loin de là les utilisateurs. Seule une écoute ou une surveillance physique permettent de mettre un individu derrière un téléphone.
- Bonjour, la classification des liens varie en fonction du type de GCO travaillé. Pour un GCO étranger qui sévit en France puis quitte le territoire, les alias seront très importants et les liens TELECOM potentiellement forts. Alors que pour la communauté des GDV, les liens téléphoniques n'auront pas grande importance, les alias seront souvent inexistantes et les adresses fluctuantes au fil des semaines. Merci pour votre travail. Cordialement,
- Bonjour, la pondération des liens est une bonne idée cependant elle ne pourra jamais être la même pour tous les dossiers. Exemple avec des voitures de location : si la société est impliquée dans le réseau la pondération sera supérieure pour les liens "utilisateur", si la société de location n'est pas dans le réseau la pondération des liens "utilisateur" devient négligeable. Pareil pour les lieux de travail, la pondération diffère si la société compte 1000 ou 4 employés.
- Bonjour, le lien le plus fort pour moi est utilisateur 1 tph vers utilisateur 2 tph
- Bonjour. Parfois l'importance du lien dépendra de sa redondance, ex des antécédents criminels (si 1 ou plusieurs arrestations ensemble pour les mêmes faits); idem pour la fréquence de fréquentation de certains lieux privés ou public (bar, coiffeur, banque, salon de massage, épicerie ...). Le caractère habituel du lien devrait jouer sur son importance. Cordialement.
- Bonne continuation
- Bonne idée de créer des entités qui sont liées indirectement entre elles.
- Ce n'est pas intuitif
- Ce questionnement mérite en effet de se poser avec les unités de terrain qui effectuent ces analyses au quotidien et pas seulement avec les services centraux.
- Certaines questions sont difficilement compréhensible
- ces intensités de lien peuvent varier suivant le groupe criminel en question
- concernant la chaîne 1, je ne suis pas d'accord. le propriétaire d'un téléphone n'en est pas forcément l'utilisateur et ne crée pas un lien entre eux mais uniquement entre les lignes téléphonique.
- Concernant la question 4, je ne vois pas l'intérêt de ce type de connexion. A l'exception de personne - personne
- Démarche très intéressante.
- En espérant avoir pu contribuer favorablement à l'étude. Bonne continuation.
- En espérant que cela puisse vous aider dans vos travaux.
- En fonction des procédures, des niveaux de force des liens peuvent être différents. L'important, est que du début à la fin d'une même procédure, l'anacrim conserve le même niveau de lien.
- Excellente étude. Bravo pour l'idée !!

- Chaque dossier et thématique sont différents. Un lien faible dans une thématique peut être fort dans un autre. C'est l'étude faite pour chaque dossier qui donne la force de chaque lien.
- Il est difficile de répondre à ce type d'enquête, car plusieurs variables sont à prendre en compte, (le type de fait, s'il s'agit d'un caln familial ou communautaire, les relations entre plusieurs communautés, ..etc..) un GCO ne fait pas non plus tous les GCO. De plus, lorsqu'un GCO est démantelé, il apprend des techniques que nous avons utilisées pour le démasquer et en tire une nouvelle tactique.
- il est difficile de déterminer la force des liens qui pour chaque dossier pourrait être différent. Néanmoins les contacts physiques et téléphoniques restent des liens forts de manière constante.
- Il est très difficile d'attribuer un niveau de force à un lien. Tout dépend du contexte de ce lien et de l'importance de celui dans le cadre de l'enquête.
- Il manque la possibilité de répondre "Force du lien inquantifiable" dans les réponses possibles. Sinon, tout dépend des investigations effectuées. Une donnée pourra être plus ou moins importante selon ce que les investigations ont préalablement mis en évidence. Toutes ces questions sont à poser aux DE finalement par rapport à leurs précédentes enquêtes.
- Il s'avère difficile de déterminer la pertinence d'un lien dans ce questionnaire en ne sachant pas sur quelles Enti têtes on travaille.
- J'ai trouvé ces chaînes de type pertinentes
- Je n'ai aucune remarque à formuler. Je vous souhaite une bonne continuation et attend avec impatience le fruit de vos travaux. Cordialement
- je pense que dans le cadre de notre travail, nous sommes déjà amenés à faire ces pondérations c'est même ce que les enquêteurs attendent de nous mais une harmonisation permettrait d'aborder certains dossiers peut être différemment
- Je pense qu'il est difficile de parler du poids du lien sans prendre en compte la valeur de la source. Tous les renseignements n'étant pas classés A1, le poids du lien sera à mon sens différent pour une même relation.
- J'espère avoir bien appréhendé le sujet et que ma réponse vous aidera...
- je trouve très intéressant et très pertinent cette notion de "force" entre entités. Bon courage à vous pour la suite.
- La démarche est intéressante. Toutefois, créer un classement en se basant sur des tendances ou des données abstraites (ex nihilo) est potentiellement subjectif. Chaque dossier est particulier et une vérité très prégnante dans l'un n'aura pas forcément son écho dans un autre dossier. Pour reprendre votre exemple de la chaîne 1 : le fait qu'une personne A possède une ligne en lien avec un numéro appartenant à B ne veut pas forcément dire que A est en lien avec B, c'est peut-être A, C ou D.
- La difficulté se trouve dans le fait que tous les liens sont potentiellement forts, les liens humains sont forts mais moins fiables que les liens matériels qui doivent être exploités
- La force d'un lien devrait pouvoir être renforcée par l'existence d'un événement particulier lié aux méfaits commis
- La majorité des réponses comportent des libellés : "lien potentiellement fort" car l'interprétation de l'importance ou non, d'un lien entre deux entités dépend de prime abord, du contexte, de l'environnement et de la connaissance du fonctionnement même des GCO par type de délinquance. Il est donc extrêmement difficile voir impossible de déterminer la pertinence d'un lien, sans ces éléments (s'agit il d'un trafic de stup ? D'un blanchiment lié à du TI ? De mules cadre cyber escroquerie ?)...
- La pondération est effectivement importante et difficile à établir. La notion qui m'interpelle est celle du lien potentiellement fort car c'est le lien le plus intéressant puisque c'est sur cette notion que nous allons développer des hypothèses à vérifier. cette notion est à mettre en évidence dans l'extraction des données, après bien sûr le travail sur les liens fort. je trouve votre travail très intéressant et pertinent. je serais intéressée de prendre connaissance de vos travaux...
- La pondération ne se limite pas au type de lien d'entités liées, mais également à la l'origine de l'information, la nature de la source (constatation enquêteur; source ouverte, réquisition, etc...), d'où la difficulté et la nécessité de votre travail ;-) Merci
- La solidité des liens évolue au cours de l'enquête. L'importance d'un lien est souvent personnel selon l'appréciation du magistrat , du DE et de l'ANACRIM qui priorisent les priorités de l'enquête. Standardiser la nature des liens pour la lecture des schémas anacrim peut permettre une compréhension plus "universelle" d'un schéma.
- la valeur de ces liens peut évoluer en fonction du domaine d'implication du GCO
- La valeur du lien entre 2 entités dépend de notre perception des données.
- le lien telecom contact telecom me parait extrêmement fort à condition que les utilisateurs réels soient identifiés sinon pour moi il serai juste fort
- Le lien telecom de la question 6 peut être majoré ou diminué en fonction de la répétition des contacts ou des types de contacts supposés (par ex on peut majorer des contacts réguliers quasi exclusifs entre deux telecom ou diminuer des contacts entre telecoms professionnels pour des commandes "honnêtes"). De la même manière, le lien professionnel entre personne et personne morale est plus ou moins fort (un dirigeant a un lien plus fort qu'un salarié ou bien même un simple démarcheur).
- L'environnement dans un dossier est important pour définir la force des liens, car en fonction du milieu d'appartenance des protagonistes (MENS, pays de l'est, stup, etc...) la force des liens et la priorisation de ces liens peut varier

- Les GCO sont protéiformes. Vouloir les personnifier me paraît dangereux, notamment pour éviter le fameux effet tunnel. L'intention reste cependant louable. Bon courage.
- Les interactions dans les GCO sont variables au regard de la structure et de l'histoire du groupe.
- les liens en téléphonie sont de moins en moins considérés du fait du nombre croissant des opérateurs de téléphonie qui ouvrent des lignes sans identités vérifiées.
- Les liens entre les entités sont une chose. Les attributs d'entités (via ANB) permettent de d'obtenir un caractère formel et véritable quant aux adresses et localisations notamment. De ce fait, les liens, aussi formels soient-ils quant aux éléments de procédure, peuvent être diminués; cela rend une lecture plus claire en modélisation. Seul "hic", l'attribut n'est pas reconnu en tant qu'entité, donc, la rigueur et le bon sens de l'enquêteur analyste restent de mise. L'outil n'est qu'une aide !
- Les liens sont définis par les éléments de procédure, ceux-ci seulement confirmeront s'ils sont "forts" ou "faibles" (observations/investigations/constatations.. ou simples renseignements)
- les liens sont difficilement définissables de façon générale/globale car variables suivant la culture/origine du GCO
- Les questions peuvent avoir une double interprétation dans la mesure où elles peuvent concerner soit la commission de l'infraction elle-même, soit l'environnement de la cible. L'intérêt des liens peut alors varier.
- L'étude des relations huaines relèvent de la sociologie et non pas de l'analyse criminelle. Chaque affaire est particulière; chaque relation humaine est particulière et unique; aucune statistique ne peut élaborer une quelconque généralité dans les relations humaines et/ou criminelles.
- L'idée de classer ces liens en fonction de leur "force" ou de leur "pertinence" est intéressante mais les cas exposés sont complètement hors contexte . Il me paraît dès lors difficile de répondre finement aux questions posées.
- L'idée de cotation pour apprécier la pertinence des liens me paraît effectivement pertinente
- L'idée est bonne en partant du principe qu'on a tous les lien, ce qui n'est jamais le cas. Et par expérience, les entités les plus liées sont souvent celles sur lesquelles l'enquête a débuté ou qui sont en position charnière, mais finalement très rarement les plus importantes dans l'organisation. Un bon chef criminel ne se salit pas les mains et reste dans l'ombre...
- L'idée est très intéressante. Cependant, je pense que l'analyste doit rester factuel. Même un système de pondération bien réalisé ne doit pas faire perdre de vue que l'analyste n'est pas un enquêteur, et qu'il doit simplement orientier ce dernier en mettant en exergue des points importants. Par rapport à tes exemples, on peut également supposer que les deux individus sont en contact téléphonique sans arrière-pensée particulière, mais qu'à l'inverse l'enquête permet de déterminer que Y employé de
- L'indice ne peut être mis automatiquement par le programme, il faut qu'il soit inséré manuellement au moment de la saisie par l'anacrim, seul à pouvoir déterminer de l'importance de la donnée liant deux entités A titre d'exemple, en prenant la procédure sur les tph encrochats, on pouvait lier les membres d'un même groupe criminel parce qu'ils communiquaient entre eux, mais celui qui était à la tête d'une des branches étaient celui qui payaient l'abonnement et non pas qui avait le plus de coms
- Ma pratique de l'analyse criminelle est moindre mais je tenais malgré tout à participer. Merci pour votre travail.
- mot
- neant
- NEANT
- néant
- (8x) Néant
- Néant.
- (4x) non
- (8x) Non
- (2x) NON
- Peut être fair un distingo entre les individus "obligés", qui vont être utilisées pour ouvrir des lignes ou louer des voitures à leurs noms pour le réseau criminel (clients mauvais payeurs, petites amies, famille, relationnel "propre" ...), et les personnes actifs dans le réseau
- Pour certaines questions, il y a une notion de subjectivité dans l'appréciation de la force que l'on peut donner au lien. Cela peut dépendre aussi du type de GCO concerné, certains étant basés sur des groupes familiaux
- Pour la question n°5 il est difficile de classer ces différents critères. Dans le cadre de GCO transnationaux le relationnel familial peut avoir autant d'importance que le lien clanique (il est même parfois lié). De même que les liens sur antécédents judiciaires sont imbriqués avec l'utilisation d'alias (alias bien plus souvent issus de saisies erronées de l'identité par les fdo que de par l'intention coupable du mis en cause d'utiliser une fausse identité)

- Pour un analyse criminel, chaque renseignement obtenu dans une enquête doit être enregistré. L'importance d'une base de données et du logiciel Analyse NoteBook permettent d'enregistrer et de sauvegarder tous ces éléments. Bon courage
- Quelque soit la force attribuée à un élément, il faut garder la même logique d'analyse. L'intensité du lien se fera t elle automatiquement selon les termes employés ?
- Questionnaire très intéressant.
- Question 6: On ne sait pas si on parle d'utilisateur/propriétaire ou juste 2 n° qui s'appellent. Un télécom s'est vague tant que nous n'avons pas la personne. Idem pour la localisation d'un télécom
- ras
- (5x) RAS
- R.A.S
- (2x) sans objet
- tout ça reste relatif
- 2 entités peuvent être reliées par plusieurs liens (ami qui est aussi son collègue de travail qui est aussi celui qui lui prête sa boiture qui) : il me semble opportun de tenir compte de la pondération du lien le plus fort/sensible mais certains autres mériteraient d'apporter quelques poids supplémentairepoinds supplmémentaires, mais pas tous

Annexe 8 Forces liantes des symboles terminaux (trinômes sémantiques) existants

Symbole terminal	Groupe sémantique	Force liante
PERSONNE-familial amical-PERSONNE	1	1
PERSONNE-en contact-PERSONNE	1	1
PERSONNE-criminel-PERSONNE	1	1
PERSONNE-est-PERSONNE	1	1
PERSONNE-fait partie-PERSONNE	1	1
PERSONNE-professionnel-PERSONNE	1	1
PERSONNE-adresse-PERSONNE	1	1
PERSONNE-utilisateur-PERSONNE	1	1
PERSONNE-échange objet ou argent-PERSONNE	1	1
PERSONNE-même nom-PERSONNE	1	0
PERSONNE-lien interdit-PERSONNE	1	0
PERSONNE-non lien-PERSONNE	1	0
PERSONNE-adresse-LIEU (ADRESSE)	2	0,8
PERSONNE-adresse-LIEU PRIVÉ	2	0,8
PERSONNE-adresse-LIEU PUBLIC	2	0,7
PERSONNE-adresse-PERSONNE MORALE	2	0,65
PERSONNE-titulaire-LIEU (ADRESSE)	2	0,3
PERSONNE-utilisateur-LIEU PRIVÉ	2	0,75
PERSONNE MORALE-titulaire-TÉLÉCOM	3	0,1
PERSONNE-titulaire-TÉLÉCOM	3	0,9
PERSONNE-utilisateur-TÉLÉCOM	3	0,9
PERSONNE-utilisateur-VÉHICULE	3	0,65
PERSONNE-titulaire-VÉHICULE	3	0,75
PERSONNE MORALE-titulaire-VÉHICULE	3	0,1 / 0,06
PERSONNE MORALE-utilisateur-VÉHICULE	3	0,09 / 0,05
PERSONNE-titulaire-OBJET	3	0,5
LIEU PUBLIC-titulaire-VÉHICULE	3	0,1
LIEU PUBLIC-titulaire-TÉLÉCOM	3	0,1 / 0,04
TÉLÉCOM-en contact-PERSONNE	4	0,7
TÉLÉCOM-en contact-TÉLÉCOM	4	0,9
PERSONNE-localisation-LIEU PUBLIC	5	0,1 / 0,55
PERSONNE-localisation-LIEU (ADRESSE)	5	0,65 / 0,6
PERSONNE-localisation-PERSONNE MORALE	5	0,2
PERSONNE-localisation-LIEU PRIVÉ	5	0,65
PERSONNE-localisation-VÉHICULE	5	0,65

Symbole terminal	Groupe sémantique	Force liante
VÉHICULE-localisation-VÉHICULE	6	0,65
VÉHICULE-localisation-PERSONNE MORALE	6	0,06
VÉHICULE-localisation-LIEU PUBLIC	6	0,08 / 0,4
VÉHICULE-localisation-LIEU (ADRESSE)	6	0,4 / 0,5
VÉHICULE-localisation-LIEU PRIVÉ	6	0,6
PERSONNE- échange objet ou argent -VÉHICULE	7	0,65
PERSONNE- échange objet ou argent -LIEU (ADRESSE)	7	0,6
PERSONNE- échange objet ou argent -LIEU PUBLIC	7	0,5 / 0,55
PERSONNE- échange objet ou argent -LIEU PRIVE	7	0,65
PERSONNE MORALE-adresse-LIEU (ADRESSE)	8	0,1
PERSONNE MORALE-titulaire-LIEU (ADRESSE)	8	0,2
OBJET-localisation-VÉHICULE	9	0,5
TÉLÉCOM-localisation-VÉHICULE	9	0,6
TÉLÉCOM-localisation-LIEU (ADRESSE)	9	0,7
OBJET-localisation-LIEU (ADRESSE)	9	0,6
PERSONNE-professionnel-PERSONNE MORALE	10	0,4
PERSONNE-titulaire-PERSONNE MORALE	10	0,5
PERSONNE-financier-PERSONNE MORALE	10	0,5
PERSONNE-professionnel-LIEU (ADRESSE)	10	0,75
PERSONNE MORALE-professionnel-LIEU (ADRESSE)	10	0,08
PERSONNE-professionnel-LIEU PUBLIC	10	0,4
OBJET-lien-VÉHICULE	11	0,75
TÉLÉCOM-lien-LIEU (ADRESSE)	11	0,7
TÉLÉCOM-lien-PERSONNE	11	0,4
VÉHICULE-lien-LIEU (ADRESSE)	11	0,65
PERSONNE-utilisateur-BANQUE	12	0,04
PERSONNE-antécédents-OBJET	12	0,04 / 0,2
PERSONNE-M.E.C.-OBJET	12	0,7
PERSONNE MORALE-autre-LIEU (ADRESSE)	13	0,3
LIEU PUBLIC-adresse-LIEU (ADRESSE)	13	0,9
PERSONNE-autre-LIEU (ADRESSE)	13	0,3
PERSONNE-autre-VÉHICULE	13	0,3
LIEU PRIVE-adresse-LIEU (ADRESSE)	13	0,9
TÉLÉCOM-autre-TÉLÉCOM	13	0,3

Annexe 9 Comparaison des taux d'erreurs obtenus en considérant des *patterns* de chemins de longueur 2 à 5 et des *patterns* de longueurs 2 à 6 (avant et après correction)

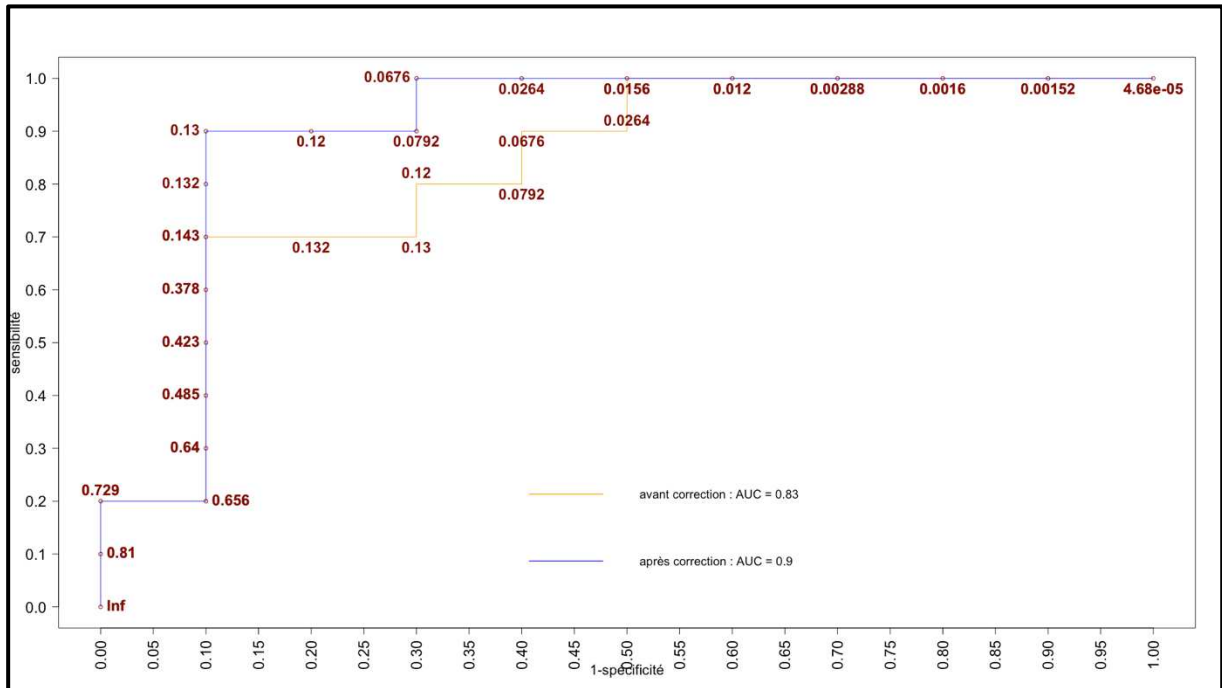


Figure 101 Courbes ROC avant et après correctifs pour un échantillon avec des longueurs de chemin de 2 à 5

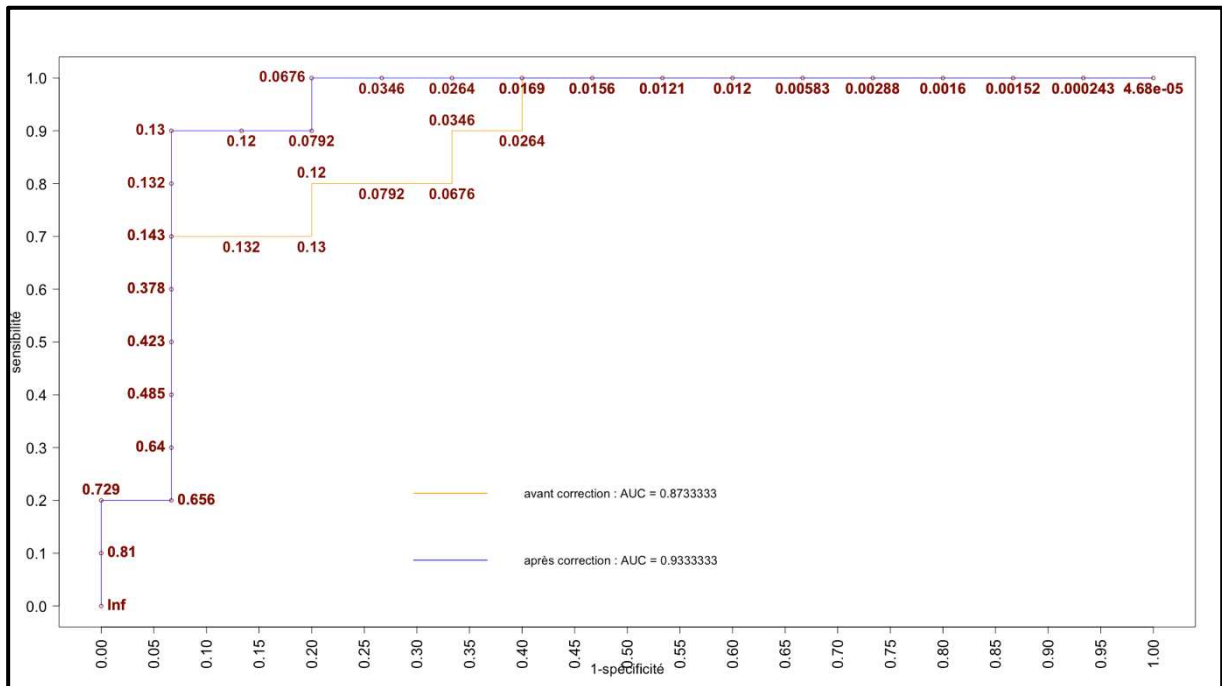


Figure 102 Courbes ROC avant et après correctifs pour un échantillon avec des longueurs de chemin de 2 à 6

En comparant les figures 101 et 102, on constate qu'à partir de la valeur seuil 0.0158, la fonction de traçage de la courbe ROC évalue des seuils plus rapprochés sur le même intervalle. Cependant, les seuils optimaux, 0.13 et 0.0676 restent inchangés, de même que la

spécificité et la sensibilité. L'aire sous la courbe est très légèrement supérieure lorsque l'on considère des chemins de longueur 2 à 6.

Pour l'échantillon étudié, l'aire sous la courbe augmente après avoir effectué les correctifs. Dans la zone où les courbes ne sont pas superposées, on peut voir que le rapport TVP/TVN est supérieur après correction qu'avant, tandis que le rapport TFP/TFN est inférieur. Les correctifs ont ainsi amélioré la précision du modèle pour l'échantillon testé, ce qui est très encourageant. Ce constat est également valable en ne considérant que 20 chemin dans l'échantillon.

Annexe 10 Détail des *patterns* de chemins évalués (échantillon) pour l'analyse des résultats de l'extraction du réseau social et la détermination de la valeur seuil de filtrage du réseau social d'enquête

N°	<i>Pattern</i>	Décision sémantique	Pond 1	Pond 2
L2.1	PERSONNE-titulaire-VÉHICULE-utilisateur-PERSONNE	Acceptation	4,85E-01	4,85E-01
L2.2	PERSONNE-adresse-LIEU ad-adresse-PERSONNE	Acceptation	6,40E-01	6,40E-01
L2.3	PERSONNE-titulaire-TEL-utilisateur-PERSONNE	Acceptation	8,10E-01	8,10E-01
L2.4	PERSONNE-utilisateur-BANQUE-utilisateur-PERSONNE	Rejet	1,60E-03	1,60E-03
L2.5	PERSONNE-localisation-VÉHICULE-utilisateur-PERSONNE	Acceptation	4,23E-01	4,23E-01
L3.1	PERSONNE-titulaire-TEL-en contact-TEL-titulaire-PERSONNE	Acceptation	7,29E-01	7,29E-01
L3.2	PERSONNE-utilisateur-VÉHICULE-titulaire-PERSONNE MORALE-professionnel-PERSONNE	Rejet	2,60E-03	1,56E-02
L3.3	PERSONNE-localisation-LIEU pu-localisation-VÉHICULE-utilisateur-PERSONNE	Acceptation	5,20E-03	1,43E-01
L3.4	PERSONNE-titulaire-LIEU ad-adresse-PERSONNE MORALE-professionnel-PERSONNE	Rejet	1,20E-02	1,20E-02
L3.5	PERSONNE-localisation-LIEU ad-lien-TEL-titulaire-PERSONNE	Acceptation	4,10E-01	3,78E-01
L4.1	PERSONNE-utilisateur-VÉHICULE-titulaire-PERSONNE MORALE-titulaire-VÉHICULE-utilisateur -PERSONNE	Rejet	4,23E-03	1,52E-03
L4.2	PERSONNE-titulaire-TEL-en contact-TEL-titulaire-LIEU pu-professionnel-PERSONNE	Acceptation	3,24E-02	1,30E-01
L4.3	PERSONNE- titulaire-TEL-en contact-TEL-en contact-TEL-titulaire-PERSONNE	Rejet	6,56E-01	6,56E-01
L4.4	PERSONNE-adresse-LIEU ad-localisation-VÉHICULE-localisation-LIEU ad-localisation-PERSONNE	Rejet	8,32E-02	1,20E-01
L4.5	PERSONNE-utilisateur-VÉHICULE-localisation-LIEU pu-localisation-VÉHICULE-utilisateur-PERSONNE	Acceptation	2,70E-03	6,76E-02
L5.1	PERSONNE-titulaire-VÉHICULE-localisation-VÉHICULE-localisation-LIEU ad-localisation-VÉHICULE-utilisateur-PERSONNE	Rejet	5,07E-02	7,92E-02
L5.2	PERSONNE-professionnel-PERSONNE MORALE-utilisateur-VÉHICULE-titulaire-PERSONNE MORALE-titulaire-VÉHICULE-utilisateur-PERSONNE	Rejet	2,34E-04	4,68E-05
L5.3	PERSONNE-utilisateur-VÉHICULE-localisation-LIEU pr-adresse-LIEU ad-localisation-VÉHICULE-titulaire-PERSONNE	Acceptation	1,05E-01	1,32E-01

N°	<i>Pattern</i>	Décision sémantique	Pond 1	Pond 2
L5.4	PERSONNE-localisation-LIEU pu-localisation-VÉHICULE- localisation-LIEU pu-localisation-VÉHICULE-titulaire- PERSONNE	Rejet	3,84E-05	2,64E-02
L5.5	PERSONNE-professionnel-PERSONNE MORALE-titulaire- VÉHICULE-localisation-LIEU pu-localisation-VÉHICULE- titulaire-PERSONNE	Rejet	1,92E-04	2,88E-03
L6.1	PERSONNE-utilisateur-TEL-localisation-VÉHICULE- localisation-LIEU pu-localisation-VÉHICULE-localisation- LIEU ad-adresse-PERSONNE	Rejet	1,11E-03	3,46E-02
L6.2	PERSONNE-adresse-LIEU ad-adresse-PERSONNE MORALE-titulaire-TEL-en contact-TEL-en contact-TEL- titulaire-PERSONNE	Rejet	5,83E-03	5,83E-03
L6.3	PERSONNE-échange objet/argent-LIEU pu-localisation - VÉHICULE-localisation-LIEU ad-localisation-VÉHICULE- localisation-LIEU pu-échange objet/argent-PERSONNE	Rejet	2,56E-04	1,21E-02
L6.4	PERSONNE-utilisateur-VÉHICULE-localisation-LIEU pu- localisation-VÉHICULE-titulaire-PERSONNE MORALE- titulaire-VÉHICULE-utilisateur-PERSONNE	Rejet	2,70E-05	2,43E-04
L6.5	PERSONNE-utilisateur-VÉHICULE-localisation-LIEU ad- localisation-VÉHICULE-localisation-LIEU pu-localisation- VÉHICULE-utilisateur-PERSONNE	Rejet	4,33E-04	1,69E-02

Abréviations employées :

LIEU pu : LIEU PUBLIC

LIEU ad : LIEU ADRESSE

LIEU pr : LIEU PRIVÉ

Annexe 11 Détail des symboles minimaux dans l'échantillon et occurrences dans le fichier de données initial pré-traité

Symboles minimaux dans l'échantillon	Occurrence dans les données
PERSONNE-utilisateur-VÉHICULE	261
VÉHICULE-localisation-LIEU (ADRESSE)	169
VÉHICULE-localisation-LIEU PUBLIC	48
PERSONNE MORALE-titulaire-VÉHICULE	33
PERSONNE-échange objet/argent-LIEU PUBLIC	4
PERSONNE-adresse-LIEU (ADRESSE)	293
PERSONNE MORALE-adresse-LIEU (ADRESSE)	97
PERSONNE MORALE-titulaire-TÉLÉCOM	11
TÉLÉCOM-en contact-TÉLÉCOM	65
PERSONNE-titulaire-TÉLÉCOM	92
PERSONNE-utilisateur-TÉLÉCOM	137
TÉLÉCOM-localisation-VÉHICULE	1
PERSONNE-professionnel-PERSONNE MORALE	92
PERSONNE-titulaire-VÉHICULE	177
VÉHICULE-localisation-LIEU PRIVÉ	12
PERSONNE MORALE-utilisateur-VÉHICULE	7
LIEU PUBLIC-titulaire-TÉLÉCOM	1
PERSONNE-localisation-LIEU (ADRESSE)	55
PERSONNE-localisation-LIEU PUBLIC	81
PERSONNE-utilisateur-BANQUE	2

Annexe 12 Scripts R écrits avec l'interface RStudio pour les générer les graphiques d'analyse du modèle

```
#Libraries
library(ggplot2)
library(dplyr)
library(MASS)
library(tidyr)
library(ROCR)
library(Hmisc)
library(pracma)

#IMPORT DATA
#data2 = 2eme set (apres correction)
data2 <- read.csv2('/Users/maudb/Desktop/These 3 - INTACTILE-IRM-
SCRC/ANALYSES/Valeur_Seuil/Echantillon_test_seuil.csv')
colnames(data2) <- c('num_ligne', 'lenght', 'chemin', 'score', 'label',
'Decision')
data2$Decision <- as.character(data2$Decision)

#1. DISTRIBUTION DES PONDERATION
#a - comparaison des distributions 'acceptation' et 'rejet' sur l
echantillon

ggplot(data2, aes(x=score)) + geom_histogram(aes(color=Decision,
fill=Decision), position = "identity", binwidth = .2) +
  scale_x_log10() +
  labs(x="Ponderations", y = "Occurrences") +
  scale_color_manual(values = c("darkgreen", "darkred")) +
  scale_fill_manual(values = c("green", "red"))

#En ajoutant la courbe de densite
ggplot(data2, aes(x=score, color=Decision, fill=Decision)) +
  #geom_histogram(aes(y = ..density..), position="identity",
binwidth = .2, alpha=0.8) +
  geom_density(alpha=0.5) +
  scale_x_log10() +
  labs(x="Ponderations", y = "Densite") +
  scale_color_manual(values = c("darkgreen", "darkred")) +
  scale_fill_manual(values = c("green", "red"))

#2. CALCUL DE TAUX D'ERREURS
#a - PREDICTION
pred2 <- prediction(data2$score, data2$label)
perf2_ROC <- performance(pred2, "tpr", "fpr")
perf2_acc <- performance(pred2, "acc")
perf2_err <- performance(pred2, "err")
perf2_prec <- performance(pred2, "prec")
perf2_mat <-performance(pred2, "mat")
perf2_auc <- performance(pred2, "auc")

#b - INDICATEURS DE PERFORMANCE
#Traitement de la classe de l objet 'perf' pour faciliter le plot
x2 <- as.numeric(perf2_ROC@x.values)
y2 <- as.numeric(perf2_ROC@y.values)
```

```

#c - PLOT
plot(x2, y2, xaxt="n", yaxt="n", type='l', xlab = '1- specificite',
ylab = 'Sensibilite', cex.lab = 1.4, col = 'blue')
lines(x2,y2, type='p', col = "darkgreen", cex=0.5)
axis(1, at = seq(0, 1, by = 0.05), las = 3, cex.axis=1.3)
axis(2, at = seq(0, 1.5, by = 0.1), las = 1, cex.axis=1.5)
text(x2, y2, labels = as.character(unlist(perf2_ROC@alpha.values)),
col = "darkgreen", font=2, cex = 1.5,
pos = c(4,4,3,4,2,2,2,2,2,2,2,1,1,2,1,1,1,1,1,1,1,1,1,1))
legend("bottomright", legend = c("Modèle : AUC = 0.9333333",
"Aléatoire : AUC = 0.5"), lwd = 1.5, cex=1.2, col = c('blue', 'black'),
lty = c(1,4), bty = "n")

#Calculer les seuils optimaux
#https://hal.archives-ouvertes.fr/hal-02870055/document
Seuils <- as.numeric(unlist(perf2_ROC@alpha.values))
#Youden -> On recherche le max de (Sensibilite + specificite)
Yd <- matrix(0, nrow=length(x2), ncol =1)

#Coin superieur gauche -> On recherche le min de (1-se)^2 + (1-sp)^2
CSG <- matrix(0, nrow=length(x2), ncol =1)

A <- as.data.frame(cbind(y2, 1-x2, x2, 1-y2, Seuils, Yd, CSG,
as.numeric(unlist(perf2_acc@y.values)),
as.numeric(unlist(perf2_err@y.values))))

colnames(A) <- c('sensibilite','spécificite','TFP','TFN','Seuils',
'Yd', 'CSG', 'précision', 'erreur globale')
#Calcul des indicateurs Youden & CSG
for (i in seq(1:nrow(A)))
{
  A$Yd[i] <- A$sensibilite[i] - A$TFP[i] +1
  A$CSG[i] <- (1-A$sensibilite[i])^2 + A$TFP[i]^2
  i=i+1
}

#Recherche des valeurs optimales
Yd_max <- max(A$Yd)
CSG_min <- min(A$CSG)
#-----

```

Annexe 13 Interprétation des indicateurs de performance utilisés en complément des courbes ROC

A l'inverse de l'aire sous la courbe ROC qui se calcule sur l'ensemble des seuils, les indicateurs suivants se calculent pour chacun des seuils évalués.

Indicateur	Calcul	Interprétation
Précision (<i>prec</i>)	$\frac{VP + VN}{P + N}$	Le taux de valeurs correctement classées
Taux d'erreur global (<i>err</i>)	$\frac{FP + FN}{P + N}$	Le taux de valeurs faussement classées
Valeur Prédictive Positive (VPP)	$\frac{VP}{VP + FP}$	Rapport des vrais positifs sur les positifs détectés par le modèle
Valeur Prédictive Négative (VPN)	$\frac{VN}{VN + FN}$	Rapport des vrais négatifs sur les négatifs détectés par le modèle
Rapport de Vraisemblance Positif (RVP)	$\frac{\text{sensibilité}}{1 - \text{spécificité}} = \frac{P(+ H_+)}{P(+ H_-)}$	Rapport de la probabilité que le test soit positif sachant que le <i>pattern</i> est accepté sur la probabilité que le test soit positif sachant que le <i>pattern</i> est rejeté
Rapport de Vraisemblance Négatif (RVN)	$\frac{\text{spécificité}}{1 - \text{sensibilité}} = \frac{P(- H_-)}{P(- H_+)}$	Rapport de la probabilité que le test soit négatif sachant que le <i>pattern</i> est rejeté sur la probabilité que le test soit négatif sachant que le <i>pattern</i> est accepté

Annexe 14 Scripts R écrits avec l'interface RStudio pour effectuer les visualisations et calculer les centralités secondaires

```
#Libraries
library(ggplot2)
library(dplyr)
library(MASS)
library(tidyr)
library(ROCR)
library(Hmisc)
library(pracma)

#IMPORT DATA
#Matrice d adjacence
adj_matrix <- read.csv2('/Users/maudb/Desktop/These 3 - INTACTILE-IRM-
SCRC/ANALYSES/INR_V5-2/SortiesTLP-2/adj_mat_INR_V5-2.csv', header=F)
adj_matrix <- as.data.frame(adj_matrix)
corresp_INR <- read.csv2('/Users/maudb/Desktop/These 3 - INTACTILE-IRM-
SCRC/ANALYSES/INR_V5-2/SortiesTLP-2/corresp_INR_V5-2.csv', header=T)
ID_tulip <- corresp_INR[,1]
#Table de centralites
centralities <- read.csv2('/Users/maudb/Desktop/These 3 - INTACTILE-
IRM-SCRC/ANALYSES/INR_V5-2/SortiesTLP-2/Centralities_INR_V5-2.csv')
centralities <- cbind(ID_tulip, centralities)
centralities$Betweenness_norm <- 2*centralities$Betweenness_norm
Closeness <- centralities$Closeness_norm
centralities$Closeness_norm <- 1 /centralities$Closeness
centralities$Closeness <- Closeness

#CREATION DES NOUVELLES COLONNES
C2 <- matrix(0,nrow=nrow(centralities), ncol =2)
colnames(C2) = c('Betweenness2', 'Degree2')
centralities <- cbind(centralities,C2)

i=1
for (i in seq(1:nrow(centralities)))
{
  j=1
  for (j in seq(1:nrow(centralities)))
  {
    centralities$Degree2[i] <- centralities$Degree2[i] +
adj_matrix[i,j]*(centralities$Degree[j] - 1)
    centralities$Betweenness2[i] <- centralities$Betweenness2[i] +
adj_matrix[i,j]*centralities$Betweenness_norm[j]
    j=j+1
  }
  centralities$Degree2[i] <-
centralities$Degree2[i]/(nrow(centralities)-1)
  i=i+1
}

#EXPORT DATA
write.csv(centralities, '/Users/maudb/Desktop/These 3 - INTACTILE-IRM-
SCRC/ANALYSES/INR_V5-2/SortiesR-2/All_centralities_INR_V5_2.csv')

#seuil 'valeurs extrêmes' (calculées avec excel)
Bx <- read.csv2('/Users/maudb/Desktop/These 3 - INTACTILE-IRM-
SCRC/ANALYSES/INR_V5-2/SortiesR-2/All_centralities_INR_V5_2.csv',
header=T)[616,4]
```

```

Bnx <- read.csv2('/Users/maudb/Desktop/These 3 - INTACTILE-IRM-
SCRC/ANALYSES/INR_V5-2/SortiesR-2/All_centralities_INR_V5_2.csv',
header=T) [616,5]
Dx <- read.csv2('/Users/maudb/Desktop/These 3 - INTACTILE-IRM-
SCRC/ANALYSES/INR_V5-2/SortiesR-2/All_centralities_INR_V5_2.csv',
header=T) [616,8]
Dnx <- read.csv2('/Users/maudb/Desktop/These 3 - INTACTILE-IRM-
SCRC/ANALYSES/INR_V5-2/SortiesR-2/All_centralities_INR_V5_2.csv',
header=T) [616,9]
B2x <- read.csv2('/Users/maudb/Desktop/These 3 - INTACTILE-IRM-
SCRC/ANALYSES/INR_V5-2/SortiesR-2/All_centralities_INR_V5_2.csv',
header=T) [616,10]
D2x <- read.csv2('/Users/maudb/Desktop/These 3 - INTACTILE-IRM-
SCRC/ANALYSES/INR_V5-2/SortiesR-2/All_centralities_INR_V5_2.csv',
header=T) [616,11]

#VISUALISATION - SCATTERPLOT
#Betweenness=f(degree)
par(mar = c(4.5, 5.5, 1, 1.5))
plot(centralities$Degree_norm, centralities$Betweenness_norm, xlab =
'Centralité de degré', ylab = 'Centralité d intermédiarité', cex = 0.7,
pch=3)
abline(h = mean(centralities$Betweenness_norm), col = "darkgreen", lty
= 5, lwd = 1)
abline(v = Dnx, col = "darkred", lty = 5, lwd = 1)
legend("top", lty = 5, legend = c( "Seuil de degrés extrêmes ", "Seuil
d'intermédiarités extrêmes"), lwd = 1.5, cex=0.8, col = c("darkred",
'darkgreen'), bty = "n")

#linear correlation
plot(centralities$Degree_norm, centralities$Betweenness_norm, xlab =
'Centralité de degré', ylab = 'Centralité d intermédiarité', cex = 0.9,
pch=3)
abline(h = mean(centralities$Betweenness_norm), col = "darkgreen", lty
= 5, lwd = 1)
abline(v = Dnx, col = "darkred", lty = 5, lwd = 1)
abline(lm(centralities$Betweenness_norm ~ centralities$Degree_norm),
col='goldenrod1', lwd =2, lty=3)
legend("top", lty = c(5, 5, 3), legend = c( "Seuil de degrés extrêmes
", "Seuil d intermédiarité \n extrêmes", 'régression linéaire \n
r2=0.4549096'), lwd = 1.5, cex=0.9,
col = c("darkred", 'darkgreen', 'goldenrod1'), bty = "n")
a1_R2 <- summary(lm(centralities$Betweenness_norm ~
centralities$Degree_norm))
a1_R2$r.squared

#Degree2=f(degree)
par(mar = c(4.5, 5.5, 1, 1.5))
plot(centralities$Degree_norm, centralities$Degree2, xlab = 'Centralité
de degré', ylab = 'Centralité secondaire de degré', cex = 0.7, pch=3)
abline(h = D2x, col = "darkgreen", lty = 5, lwd = 1)
abline(v = Dnx, col = "darkred", lty = 5, lwd = 1)
legend("right", lty = 5, legend = c( "Seuil de degrés extrêmes ", "Seuil
de degrés secondaires \n extrêmes"), lwd = 1.5, cex=0.7, col =
c("darkred", 'darkgreen'), bty = "n")

#Linear correlation
par(mar = c(4.5, 5.5, 1, 1.5))
plot(centralities$Degree_norm, centralities$Degree2, xlab = 'Centralité
de degré', ylab = 'Centralité secondaire de degré', cex = 0.9, pch=3)
abline(h = D2x, col = "darkgreen", lty = 5, lwd = 1)

```



```

abline(v = Dnx, col = "darkred", lty = 5, lwd = 1)
abline(lm(centralities$Degree2 ~ centralities$Degree_norm),
col='goldenrod1', lwd =2)
legend(x=0.03, y=3.25, lty = c(5, 5, 1), legend = c( "Seuil de degrés
extrêmes ", "Seuil de degrés secondaires \n extrêmes", 'régression
linéaire \n r2 = 0.8258025'), lwd = 1.5, cex=0.9,
      col = c("darkred", 'darkgreen', 'goldenrod1'), bty = "n")
a1_R2 <- summary(lm(centralities$Degree2 ~ centralities$Degree_norm))
a1_R2$r.squared

#Betweenness2=f(degree)
par(mar = c(4.5, 5.5, 1, 1.5))
plot(centralities$Degree_norm, centralities$Betweenness2, xlab =
'Centralité de degré', ylab = 'Centralité secondaire \n d
intermédiarité', cex = 0.7, pch=3)
abline(v = Dnx, col = "darkgreen", lty = 5, lwd = 1)
abline(h = B2x, col = "darkred", lty = 5, lwd = 1)
legend("bottomright", lty = 5, legend = c( "Seuil de degrés
extrêmes", "Seuil d'intermédiarités \n secondaires extrêmes"), lwd =
1.5, cex=0.7, col = c("darkred", 'darkgreen'), bty = "n")

#linear correlation
plot(centralities$Degree_norm, centralities$Betweenness2, xlab =
'Centralité de degré', ylab = 'Centralités secondaires \n d
intermédiarité', cex = 0.9, pch=3)
abline(h = mean(centralities$Betweenness2), col = "darkgreen", lty = 5,
lwd = 1)
abline(v = Dnx, col = "darkred", lty = 5, lwd = 1)
abline(lm(centralities$Betweenness2 ~ centralities$Degree_norm),
col='goldenrod1', lwd =2, lty=3)
legend(x=0.09, y=0.3, lty = c(5, 5, 3), legend = c( "Seuil de degrés
extrêmes ", "Seuil d'intermédiarités \n secondaires extrêmes",
'régression linéaire \n r2=0.521672'), lwd = 1.5, cex=0.9,
      col = c("darkred", 'darkgreen', 'goldenrod1'), bty = "n")
a1_R2 <- summary(lm(centralities$Betweenness2 ~
centralities$Degree_norm))
a1_R2$r.squared

#Betweenness2=f(deg2)
par(mar = c(4.5, 5.5, 1, 1.5))
plot(centralities$Degree2, centralities$Betweenness2, xlab =
'Centralité de degré', ylab = 'Centralité secondaire \n d
intermédiarité', cex = 0.7, pch=3)
abline(v = D2x, col = "darkgreen", lty = 5, lwd = 1)
abline(h = B2x, col = "darkred", lty = 5, lwd = 1)
legend("bottomright", lty = 5, legend = c( "Seuil de degrés secondaires
\n extrêmes", "Seuil d'intermédiarités \n secondaires extrêmes"), lwd =
1.5, cex=0.7, col = c("darkred", 'darkgreen'), bty = "n")

#Betweenness2=f(betw)
par(mar = c(4.5, 5.5, 1, 1.5))
plot(centralities$Betweenness, centralities$Betweenness2, xlab =
'Centralité d'intermédiarité', ylab = 'Centralité secondaire \n d
intermédiarité', cex = 0.7, pch=3)
abline(v = Bnx, col = "darkgreen", lty = 5, lwd = 1)
abline(h = B2x, col = "darkred", lty = 5, lwd = 1)
legend("bottomright", lty = 5, legend = c( "Seuil d'intermédiarités
extrêmes", "Seuil d'intermédiarités \n secondaires extrêmes"), lwd =
1.5, cex=0.7, col = c("darkred", 'darkgreen'), bty = "n")

#linear correlation

```

```

par(mar = c(4.5, 5.5, 1, 1.5))
plot(centralities$Betweenness_norm, centralities$Betweenness2, xlab =
'Centralité d intermédiarité', ylab = 'Centralité secondaire d
intermédiarité', cex = 0.9, pch=3)
abline(h = D2x, col = "darkgreen", lty = 5, lwd = 1)
abline(v = Dnx, col = "darkred", lty = 5, lwd = 1)
abline(lm(centralities$Betweenness2 ~ centralities$Betweenness_norm),
col='goldenrod1', lwd =2)
legend("right", lty = c(5, 5, 1), legend = c( "Seuil d intermédiarité
extrêmes ", "Seuil d intermédiarité secondaires \n extrêmes",
'«régression linéaire» \n r2 = 0.1340611'), lwd = 1.5, cex=0.9,
col = c("darkred", 'darkgreen', 'goldenrod1'), bty = "n")
a1_R2 <- summary(lm(centralities$Betweenness2 ~
centralities$Betweenness))
a1_R2$r.squared

#En distinguant les MeC
#Betweenness=f(degree)
colors <- c("blue","red")
mycol <- rgb(0, 0, 255, max = 255, alpha = 0, names = "blue_transp")
colors2 <-c(mycol, "red")
par(mar = c(4.5, 5.5, 1, 1.5))
plot(centralities$Degree_norm, centralities$Betweenness_norm, xlab =
'Centralité de degré',
ylab = 'Centralité d inermédiarité', cex = 0.7, col
=colors[as.factor(centralities$MeC)], pch=3)
text(centralities$Degree_norm, centralities$Betweenness_norm, labels =
as.character(centralities$viewLabel),
col = colors2[as.factor(centralities$MeC)], font=2, cex = 0.8, pos
= 3)
legend(x=0.035 ,y= 0.135, lty = 0, pch = 3, legend = c( "", "Prévenus
\n(ID d anonymisation)"), lwd = 1.5, cex=0.9,
col = c(mycol, 'red'), bty = "n")
abline(h = Bnx, col = "darkgreen", lty = 5, lwd = 1)
abline(v = Dnx, col = "darkred", lty = 5, lwd = 1)
legend("top", lty = 5, legend = c( "Seuil de degrés extrêmes ", "Seuil
d'intermédiarités extrêmes"), lwd = 1.5, cex=0.9,
col = c("darkred", 'darkgreen'), bty = "n")

#Degree2=f(degree)
#linear correlation
par(mar = c(4.5, 5.5, 1, 1.5))
plot(centralities$Degree_norm, centralities$Degree2, xlab = 'Centralité
de degré', ylab = 'Centralité secondaire de degré',
cex = 0.7, col =colors[centralities$MeC], pch=3)
text(centralities$Degree_norm, centralities$Degree2, labels =
as.character(centralities$viewLabel),
col = colors2[centralities$MeC], font=2, cex = 0.8, pos = 3)
legend(x=28 , y=1800 , lty = 0, pch = 3, legend = c( "", "Mis en
cause"), lwd = 1.5, cex=0.5,
col = c(mycol, 'red'), bty = "n")
abline(h = D2x, col = "darkgreen", lty = 5, lwd = 1)
abline(v = Dnx, col = "darkred", lty = 5, lwd = 1)
abline(lm(centralities$Degree2 ~ centralities$Degree_norm),
col='goldenrod1', lwd =2)
legend("right", lty = c(5, 5, 1), legend = c( "Seuil de degrés extrêmes
", "Seuil de degrés secondaires \n extrêmes", 'régression linéaire'),
lwd = 1.5, cex=0.7,
col = c("darkred", 'darkgreen', 'goldenrod1'), bty = "n")

```

```

#Betweenness2=f(degree)
par(mar = c(4.5, 5.5, 1, 1.5))
plot(centralities$Degree_norm, centralities$Betweenness2, xlab =
'Centralité de degré', ylab = 'Centralité secondaire \n d
intermédiarité',
      cex = 0.7, col =colors[as.factor(centralities$MeC)], pch=3)
text(centralities$Degree_norm, centralities$Betweenness2, labels =
as.character(centralities$viewLabel),
      col = colors2[as.factor(centralities$MeC)], font=2, cex = 0.8, pos
= 3)
abline(h = B2x, col = "darkgreen", lty = 5, lwd = 1)
abline(v = Dnx, col = "darkred", lty = 5, lwd = 1)
legend(x=0.09, y=0.2, lty = c(5,5), legend = c("Seuil de degrés
extrêmes","Seuil d'intermédiarités \nsecondaires extrêmes"), lwd = 1,
cex=0.8,
      col = c("darkred", 'darkgreen'), bty = "n")
legend(x=0.0923, y=0.15, lty = 0, pch = 3, legend = c("", "Prévenus
\n(ID d anonymisation)"), lwd = 1, cex=0.8, col = c(mycol, 'red'), bty
= "n")

#Betweenness2=f(deg2)
par(mar = c(4.5, 5.5, 1, 1.5))
plot(centralities$Degree2, centralities$Betweenness2, xlab =
'Centralité de degré', ylab = 'Centralité secondaire \n d
intermédiarité', cex = 0.7, col =colors[as.factor(centralities$MeC)],
pch=3)
text(centralities$Degree_norm, centralities$Betweenness2, labels =
as.character(centralities$ID_anonym),
      col = colors2[as.factor(centralities$MeC)], font=2, cex = 0.8, pos
= 3)

abline(v = D2x, col = "darkgreen", lty = 5, lwd = 1)
abline(h = B2x, col = "darkred", lty = 5, lwd = 1)
legend("bottomright", lty = 5, legend = c( "Seuil de degrés secondaires
\n extrêmes","Seuil d'intermédiarités \n secondaires extrêmes"), lwd =
1.5, cex=0.5, col = c("darkred", 'darkgreen'), bty = "n")

#DISTRIBUTION DES VALEURS DE CENTRALITÉ
#1 - Degree
#a - Densité
par(mar = c(4.5, 5.5, 1, 1.5))
plot(density(centralities$Degree), main='',
xlim=c(min(centralities$Degree),max(centralities$Degree)), xlab =
'Nombre de connexions directes',
      ylab = 'densite', type="l", col="blue", yaxt="n")

#b - Boxplot
par(mar = c(1, 5, 1, 1.5))
boxplot(centralities$Degree, ylab = 'Nombre de connexions directes',
col = "blue", outline = T)
segments(x0 = 0.8, y0 = mean(centralities$Degree),
x1 = 1.2, y1 = mean(centralities$Degree),
col = "red", lwd = 2, lty = 5)
legend("topright", lty = c(5, 1), lwd = 2, legend = c("moyenne",
"médiane"), col = c("red", "black"), bty = "n")

#2 - Degree_norm
#a - Densité

```

```

par(mar = c(4.5, 5.5, 1, 1.5))
plot(density(centralities$Degree_norm), main='',
xlim=c(min(centralities$Degree_norm),max(centralities$Degree_norm)),
      xlab = 'Centralité de degré standardisée', ylab = 'densite',
      type="l", col="blue")

#b - Boxplot
par(mar = c(1, 5, 1, 1.5))
boxplot(centralities$Degree_norm, ylab = 'Centralité de degré
standardisée', col = "blue") #sans outliers ", outline = F"
segments(x0 = 0.8, y0 = mean(centralities$Degree_norm),
          x1 = 1.2, y1 = mean(centralities$Degree_norm),
          col = "red", lwd = 2, lty = 5)
legend("topright", lty = c(5, 1), lwd = 2, legend = c("moyenne",
"médiane"), col = c("red", "black"), bty = "n")

#3 - Betweenness
#a - Densité
par(mar = c(4.5, 5, 1, 1.5))
plot(density(centralities$Betweenness), main='',
xlim=c(min(centralities$Betweenness), max(centralities$Betweenness)),
      xlab = 'Centralité d intermédierité',
      ylab = 'densité', type = "l", col = "orange")

#b - Boxplot
par(mar = c(1, 5, 1, 1.5))
boxplot(centralities$Betweenness, ylab = 'Centralité d intermédierité',
col = "orange", outline = F)
segments(x0 = 0.8, y0 = mean(centralities$Betweenness),
          x1 = 1.2, y1 = mean(centralities$Betweenness),
          col = "red", lwd = 2, lty = 5)
legend("topright", lty = 1, lwd = 2, legend = "médiane", col =
"black", bty = "n")

#4 - Betweenness_norm
#a - Densité
par(mar = c(4.5, 5, 1, 1.5))
plot(density(centralities$Betweenness_norm*100), main='', type = "l",
col = "orange",
      xlab = 'Chances d appartenance à \n un chemin géodésique', ylab =
'densité', xlim =c(min(centralities$Betweenness_norm*100),
max(centralities$Betweenness_norm*100)),
      yaxt = "n")

#b - Boxplot
par(mar = c(1, 5, 2, 1.5))
boxplot(centralities$Betweenness_norm*100, ylab = 'Chances d
appartenance à \n un chemin géodésique', col = "orange", outline = F)
segments(x0 = 0.8, y0 = mean(centralities$Betweenness_norm*100),
          x1 = 1.2, y1 = mean(centralities$Betweenness_norm*100),
          col = "red", lwd = 2, lty = 5)
legend("topright", lty = 1, lwd = 2, legend = "médiane", col =
"black", bty = "n")

#5 - Degree2
#a - Densité
par(mar = c(4.5, 5.5, 1, 1.5))

```

```

plot(density(centralities$Degree2), main='',
xlim=c(min(centralities$Degree2),max(centralities$Degree2)),
      xlab = 'Centralité secondaire de degré', ylab = 'densité',
type="l", col="green")

#b - Boxplot
par(mar = c(1, 5, 1, 1.5))
boxplot(centralities$Degree2, ylab = 'Centralité secondaire de degré',
col = "green") #sans outliers ", outline=F"
segments(x0 = 0.8, y0 = mean(centralities$Degree2),
      x1 = 1.2, y1 = mean(centralities$Degree2),
      col = "red", lwd = 2, lty = 5)
legend("topright", lty = c(5, 1), lwd = 2, legend = c("moyenne",
"médiane"), col = c("red", "black"), bty = "n")

#6 - Betweenness2
#a - Densité
par(mar = c(4.5, 5, 1, 1.5))
plot(density(centralities$Betweenness2), main='',
xlim=c(min(centralities$Betweenness2), max(centralities$Betweenness2)),
      xlab = 'Centralité secondaire d intermédiarité', ylab = 'densité',
type = "l", col = "purple")

#b - Boxplot
par(mar = c(1, 5, 1, 1.5))
boxplot(centralities$Betweenness2, ylab = 'Centralité d intermédiarité
standardisée', col = "purple", outline = F)
segments(x0 = 0.8, y0 = mean(centralities$Betweenness2),
      x1 = 1.2, y1 = mean(centralities$Betweenness2),
      col = "red", lwd = 2, lty = 5)
legend("topright", lty = c(5, 1), lwd = 2, legend = c("moyenne",
"médiane"), col = c("red", "black"), bty = "n")

#Comparaison des centralités
#Creation des dataframe
#1 - Standardisées
D_D2 <- cbind(centralities$Degree_norm, centralities$Degree2)
colnames(D_D2) <- c("Degré standardisé", "Degré secondaire")
#2 - Secondaires
B_B2 <- cbind(centralities$Betweenness_norm, centralities$Betweenness2)
colnames(B_B2) <- c("Intermédiarité standardisée", "Intermédiarité \n
secondaire")

#Boxplots
#1 - Standardisées
par(mar = c(4.5, 5.5, 1, 1.5))
boxplot(D_D2, col = c("blue", "green"), ylab = "Centralités",
outline=F)
invisible(lapply(1:ncol(D_D2),
      function(i) segments(x0 = i - 0.4, y0 = mean(D_D2[,
i]), x1 = i + 0.4, y1 = mean(D_D2[, i]),
      col = "red", lty = 5 , lwd = 2)))
legend("topleft", lty = c(5, 1), lwd = 2, legend = c("moyenne",
"médiane"), col = c("red", "black"), cex = 0.9, bty = "n")

#2 - Secondaires
par(mar = c(4.5, 5.5, 1, 1.5))
boxplot(B_B2, col = c("orange", "purple"), ylab = "Centralités",
outline=F)
invisible(lapply(1:ncol(B_B2),

```

```
function(i) segments(x0 = i - 0.4, y0 = mean(B_B2[,  
i]), x1 = i + 0.4, y1 = mean(B_B2[, i]),  
col = "red", lty = 5 , lwd = 2))  
legend("topleft", lty = c(5, 1), lwd = 2, legend = c("moyenne",  
"médiane"), col = c("red", "black"), cex = 0.9, bty = "n")
```

Annexe 15 Centralité de degré des 62 acteurs du réseau pour lesquels cet indicateur prend des valeurs extrêmes (par rapport à l'ensemble des 615 acteurs)

ID_anonym	Degree	Degree_norm
1	83	13,52%
250	78	12,70%
348	74	12,05%
110	68	11,07%
345	62	10,10%
344	62	10,10%
402	50	8,14%
1115	49	7,98%
288	48	7,82%
113	45	7,33%
272	43	7,00%
265	39	6,35%
259	39	6,35%
319	39	6,35%
358	39	6,35%
292	38	6,19%
318	38	6,19%
356	38	6,19%
1684	37	6,03%
298	37	6,03%
379	37	6,03%
243	34	5,54%
244	34	5,54%
252	32	5,21%
238	31	5,05%
239	31	5,05%
240	31	5,05%
241	31	5,05%
242	31	5,05%
245	31	5,05%
246	31	5,05%
247	31	5,05%
248	31	5,05%
357	30	4,89%
303	30	4,89%
293	30	4,89%
354	30	4,89%
27	29	4,72%
43	29	4,72%
48	29	4,72%
324	26	4,23%
112	25	4,07%
352	25	4,07%
59	24	3,91%
337	23	3,75%
347	23	3,75%
384	23	3,75%
56	22	3,58%
57	22	3,58%
58	22	3,58%
109	22	3,58%
521	22	3,58%
2	21	3,42%
11	21	3,42%
523	21	3,42%
9	20	3,26%
23	20	3,26%
346	20	3,26%
107	18	2,93%
393	18	2,93%
511	18	2,93%
45	17	2,77%

Annexe 16 Centralité d'intermédiation standardisée (pourcentage d'appartenance à l'un des chemins géodésiques existants entre deux nœuds quelconques) des 146 acteurs du réseau pour lesquels cet indicateur prend des valeurs extrêmes (par rapport à l'ensemble des 615 acteurs)

ID_anonym	Betweenness	Betweenness_norm
1	24848,600	13,20%
348	18493,700	9,83%
1115	14208,600	7,55%
250	14012,400	7,45%
110	9944,750	5,28%
2	6990,150	3,71%
345	6479,280	3,44%
344	6479,280	3,44%
402	6069,660	3,23%
9	4150,080	2,21%
15	3836,460	2,04%
45	3553,330	1,89%
27	3439,800	1,83%
11	3397,670	1,81%
16	3241,500	1,72%
115	2688,500	1,43%
23	2396,980	1,27%
379	2383,560	1,27%
243	2222,090	1,18%
48	2199,150	1,17%
43	2191,460	1,16%
197	2164,110	1,15%
354	1987,270	1,06%
288	1917,010	1,02%
81	1892,860	1,01%
427	1886,830	1,00%
272	1839,530	0,98%
195	1825,950	0,97%
276	1820,000	0,97%
530	1814,000	0,96%
89	1810,000	0,96%
8	1745,000	0,93%
337	1545,560	0,82%
397	1458,000	0,77%
255	1458,000	0,77%
14	1416,000	0,75%
451	1392,000	0,74%
410	1295,330	0,69%

ID_anonym	Betweenness	Betweenness_norm
113	1276,460	0,68%
435	1229,000	0,65%
218	1229,000	0,65%
228	1095,000	0,58%
1505	1095,000	0,58%
273	1095,000	0,58%
342	1092,500	0,58%
319	1067,670	0,57%
498	1044,000	0,55%
19	996,460	0,53%
411	885,000	0,47%
421	884,000	0,47%
439	884,000	0,47%
485	882,000	0,47%
420	879,000	0,47%
393	872,015	0,46%
392	858,261	0,46%
347	851,915	0,45%
275	841,918	0,45%
398	837,672	0,45%
112	827,286	0,44%
358	825,098	0,44%
307	755,298	0,40%
271	744,971	0,40%
256	732,750	0,39%
301	732,701	0,39%
52	731,000	0,39%
394	731,000	0,39%
872	730,000	0,39%
363	730,000	0,39%
353	730,000	0,39%
198	727,790	0,39%
244	714,747	0,38%
418	710,000	0,38%
216	704,000	0,37%
407	704,000	0,37%
318	701,668	0,37%
356	701,668	0,37%
292	685,921	0,36%
201	681,321	0,36%
298	649,194	0,34%
324	644,432	0,34%
415	638,833	0,34%

ID_anonym	Betweenness	Betweenness_norm
209	617,500	0,33%
384	598,536	0,32%
168	577,250	0,31%
352	565,920	0,30%
346	554,210	0,29%
1326	544,500	0,29%
1327	544,500	0,29%
152	541,367	0,29%
151	540,500	0,29%
430	534,000	0,28%
417	533,000	0,28%
59	463,286	0,25%
47	450,730	0,24%
117	404,917	0,22%
29	397,784	0,21%
279	378,735	0,20%
280	378,735	0,20%
533	366,000	0,19%
534	366,000	0,19%
515	366,000	0,19%
388	366,000	0,19%
149	366,000	0,19%
251	366,000	0,19%
295	366,000	0,19%
299	366,000	0,19%
320	366,000	0,19%
336	366,000	0,19%
362	366,000	0,19%
364	366,000	0,19%
438	357,000	0,19%
422	357,000	0,19%
475	357,000	0,19%
426	357,000	0,19%
102	356,000	0,19%
214	356,000	0,19%
472	356,000	0,19%
20	266,460	0,14%
121	266,460	0,14%
213	264,000	0,14%
441	264,000	0,14%
371	258,477	0,14%
265	228,376	0,12%
303	196,488	0,10%

ID_anonym	Betweenness	Betweenness_norm
266	182,500	0,10%
267	182,500	0,10%
338	182,500	0,10%
339	182,500	0,10%
161	180,200	0,10%
154	180,000	0,10%
104	179,000	0,10%
317	179,000	0,10%
18	179,000	0,10%
440	179,000	0,10%
150	179,000	0,10%
481	179,000	0,10%
107	143,000	0,08%
3	132,389	0,07%
46	124,638	0,07%
260	124,072	0,07%
447	124,000	0,07%
93	121,295	0,06%
98	121,295	0,06%
323	116,917	0,06%
259	112,289	0,06%
521	111,787	0,06%

Annexe 17 Liste des acteurs des quatre profils de centralité d'intérêt

Profil 3	Profil 4	Profil 7	Profil 8
1	16	1	3
2	872	27	29
9	3	43	32
11	14	48	323
23	15	56	392
27	52	57	
43	18	58	
45	19	59	
46	20	109	
48	1505	112	
59	29	113	
112	47	1684	
113	81	110	
110	89	265	
1115	93	238	
265	98	239	
243	102	240	
244	104	241	
250	115	242	
345	117	243	
344	121	244	
298	150	245	
259	151	246	
272	152	247	
303	154	248	
107	161	250	
288	168	345	
292	534	344	
318	195	357	
319	201	298	
324	197	252	
337	198	259	
346	266	272	
347	267	303	
348	260	288	
354	209	292	
352	149	293	
356	213	318	
358	214	319	
379	216	324	
384	218	337	

Profil 3	Profil 4	Profil 7	Profil 8
393	228	346	
402	251	347	
521	255	348	
	256	354	
	271	352	
	273	356	
	275	358	
	276	379	
	279	384	
	280	393	
	8	402	
	295	521	
	299	523	
	301		
	307		
	317		
	320		
	323		
	336		
	338		
	342		
	339		
	397		
	1326		
	1327		
	515		
	362		
	530		
	353		
	371		
	398		
	363		
	364		
	394		
	533		
	388		
	392		
	407		
	410		
	411		
	415		
	417		
	418		

Profil 3	Profil 4	Profil 7	Profil 8
	420		
	421		
	422		
	426		
	427		
	430		
	435		
	451		
	438		
	439		
	440		
	441		
	447		
	472		
	475		
	481		
	485		
	498		

Table des matières

Remerciements - 2 -

INTRODUCTION - 6 -

Une histoire de la police guidée par le renseignement : les exemples de la France, du Royaume-Uni et des États-Unis. - 6 -

Historique de la recherche : l'élaboration du sujet. - 15 -

Sommaire - 21 -

CHAPITRE I GENÈSE DU PROJET ET CONTEXTUALISATION DU SUJET - 22 -

I.1 Historique de la recherche : du projet au doctorat...... - 22 -

I.2 Les acteurs du projet et la construction du sujet...... - 26 -

a) L'Institut de Recherche Montesquieu (IRM) : le laboratoire de recherche académique et la direction de thèse - 27 -

b) Intactile Design, l'entreprise d'accueil - 28 -

c) Le Service central du Renseignement criminel de la Gendarmerie nationale (SCRC) : l'apport des données et de l'expertise métier - 29 -

I.3 Définition des termes du sujet et des théories mises en application...... - 30 -

I.4 Le cycle du renseignement : de la donnée brute au renseignement...... - 36 -

CHAPITRE II ÉTAT D'AVANCEMENT DES CONNAISSANCES : DES SCIENCES SOCIALES À LA MODÉLISATION - 41 -

II.1 Théories issues des sciences sociales...... - 41 -

a) Le concept d'espace social - 41 -

1. Le capital économique - 46 -

2. Le capital culturel - 46 -

3. Le capital social - 47 -

4. Le capital symbolique - 47 -

b) Développement de la sociologie de la déviance - 49 -

1. Le déterminisme - 50 -

2. L'interactionnisme (École de Chicago) - 51 -

3. L'individualisme - 54 -

c) Le crime comme occupation - 58 -

d) Les capitaux dans l'espace social criminel - 60 -

II.2 Principes d'analyse de réseaux sociaux...... - 63 -

a) Analyse de réseaux - généralités - 63 -

b) Contributions de la théorie des graphes en ARS - 64 -

1. Éléments de base en théorie des graphes : définitions et formalismes - 65 -

2. Application de la théorie des graphes en ARS - 67 -

c) Mesures de centralité - 68 -

1. La centralité de degré (*degree centrality*) - 68 -

2. La centralité de proximité (*closeness centrality*) - 69 -

3. La centralité d'intermédiarité (*betweenness centrality*) - 70 -

4. La centralité de vecteur propre (*eigenvector centrality*) - 71 -

II.3 Réseaux criminels : modélisations et connaissances...... - 73 -

a) Visualisation relationnelle - 73 -

b) Réseau comme structure et réseau comme acteur - 76 -

II.4 Éléments de méthodes analytiques appliquées pour la modélisation de réseaux. - 78 -

a) Transformations de modèles : application de la théorie des graphes - 79 -

b) Traitement automatisé du langage naturel : principe général de grammaire formelle - 80 -

CHAPITRE III	CAS D'USAGE ET DONNÉES D'ÉTUDE : PRÉSENTATION D'UNE ENQUÊTE D'ENVERGURE ET STRUCTURATION DES DONNÉES	82 -
III.1	<i>Accès aux données : spécificités administratives et pratiques</i>	82 -
a)	Aspects administratifs	82 -
b)	Aspects pratiques	83 -
III.2	<i>Description des données à disposition : forme, structure et contenu</i>	84 -
a)	Résumé de l'enquête	84 -
b)	Procès-verbaux (PV) de procédure : la donnée non structurée	86 -
c)	Matrices d'analyse : la donnée structurée	88 -
1.	Description de la structure matricielle	88 -
2.	Description des entités et de leurs attributs	90 -
3.	Relations décrites et informations associées	93 -
III.3	<i>Saisie des données : choix du matériel et de la méthode de saisie</i>	95 -
III.4	<i>Principaux outils de travail</i>	97 -
a)	L'outil <i>Stemic</i> développé et commercialisé par Intactile Design	97 -
b)	L'outil <i>Tulip</i> développé par le LaBRI	98 -
c)	Traitements de données : <i>Excel</i> et <i>RStudio</i>	99 -
CHAPITRE IV	DU PRÉ-TRAITEMENT DES DONNÉES BRUTES AU GRAPHE MULTIVARIÉ	100 -
IV.1	<i>Préambules de l'analyse : pré-traitements et traitements des données</i>	102 -
a)	Uniformisation structurelle des données, préambule nécessaire à l'extraction automatisée de connaissance.....	103 -
b)	Nettoyage et appropriation du contenu	105 -
c)	Standardisation du contenu par la création d'ontologies sémantiques : préparation aux traitements automatisés.....	108 -
IV.2	<i>Visualisations graphiques</i>	111 -
a)	Création du graphe relationnel <i>n-mode</i>	111 -
b)	Détection graphique d'anomalies dans les données	115 -
CHAPITRE V	MÉTHODOLOGIE GÉNÉRALE : DU GRAPHE MULTIVARIÉ AU RÉSEAU SOCIAL	117 -
V.1	<i>Extraction sémantique du réseau social depuis le graphe multivarié : approche par la science des réseaux</i> - 117 -	
a)	Nécessité	117 -
b)	Méthodologie de transformation : inspiration issue des concepts de grammaires formelles.....	119 -
c)	Définition de la notion de « force liante » d'un symbole	123 -
d)	Pré-traitements graphiques : réduction de l'espace de recherche	125 -
e)	Traitements du graphe : réduction de la complexité	130 -
f)	Traitements des sous-graphes « personne-personne » <i>path_{ij}</i> : obtention des liens directs pour les cas les plus complexes.....	139 -
V.2	<i>Paramétrage de l'algorithme d'extraction</i>	145 -
a)	Attribution des valeurs de force liante des symboles terminaux.....	145 -
b)	Règles de réécriture en série et en dérivation	156 -
c)	Le réseau social pondéré <i>Social_network</i>	159 -
CHAPITRE VI	DISCUSSION DE LA MÉTHODE D'EXTRACTION ET DÉTERMINATION DU SEUIL DE FILTRAGE	163 -
VI.1	<i>Analyse des résultats de l'extraction du graphe social <i>Social_network</i> depuis le graphe multivarié <i>data0</i></i>	163 -
a)	Analyse des pondérations des liens du graphe <i>Social_network</i>	163 -
b)	Optimisation des paramètres.....	166 -
1.	Pour les « LIEUX PUBLICS »	168 -
2.	Pour les « PERSONNES MORALES ».....	168 -
3.	Pour les chaînes téléphoniques :.....	169 -
4.	Informations manquantes.....	169 -

VI.2	<i>Étude des performances et détermination du seuil de filtrage</i>	- 171 -
a)	Analyse des performances du modèle	- 172 -
b)	Détermination du seuil de filtrage des pondérations	- 177 -
VI.3	<i>Discussion et élaboration de critères d'optimisation</i>	- 183 -
a)	Limites du modèle.....	- 183 -
b)	Éléments de solutions : optimisation de l'algorithme.....	- 186 -
c)	Critères de collecte et de saisie de la donnée par les enquêteurs et analystes.....	- 189 -
CHAPITRE VII	ANALYSE DE RÉSEAUX SOCIAUX : APPLICATION AU SOCIOGRAMME <i>INR</i>	- 192 -
VII.1	<i>Propriétés du sociogramme INR</i>	- 193 -
a)	Traitements préliminaires à l'analyse	- 193 -
b)	Description du graphe final : propriétés structurelles	- 195 -
VII.2	<i>Analyse du réseau social : propriétés sociométriques des nœuds</i>	- 198 -
a)	Centralité de degré.....	- 199 -
b)	Centralité de proximité	- 203 -
c)	Centralité d'intermédiarité	- 207 -
VII.3	<i>Centralités secondaires</i>	- 212 -
a)	Centralité secondaire de degré	- 213 -
b)	Centralité secondaire d'intermédiarité.....	- 219 -
VII.4	<i>Détection des positions stratégiques</i>	- 222 -
VII.5	<i>Comparaison des résultats analytiques et des conclusions de l'enquête judiciaire</i>	- 228 -
CHAPITRE VIII	INTERPRÉTATIONS EN TERMES D'ESPACE ET DE CHAMP CRIMINELS : LE COUP DE TONNERRE DU 10 NOVEMBRE	- 233 -
CONCLUSION	- 240 -	
BIBLIOGRAPHIE	- 245 -	
Table des tableaux et figures	- 264 -	
ANNEXES	- 276 -	
<i>Table des annexes</i>	- 276 -	
<i>Annexe 1 Table des sigles et acronymes employés</i>	- 278 -	
<i>Annexe 2 "Peel's nine police principles"</i>	- 280 -	
<i>Annexe 3 Ontologie de classification sémantique des liens</i>	- 281 -	
<i>Annexe 4 Circulaire ministérielle n°30053 du 3 septembre 2019 relative à l'export de données à destination de la recherche scientifique</i>	- 302 -	
<i>Annexe 5 Liste, occurrences et fréquence des symboles terminaux existant dans les données prétraitées</i> -	307 -	
<i>Annexe 6 Questionnaire soumis aux analystes de la Gendarmerie nationale pour appuyer l'attribution des forces liantes des symboles terminaux</i>	- 309 -	
<i>Annexe 7 Résultats du questionnaire</i>	- 314 -	
<i>Annexe 8 Forces liantes des symboles terminaux (trinômes sémantiques) existants</i>	- 324 -	
<i>Annexe 9 Comparaison des taux d'erreurs obtenus en considérant des patterns de chemins de longueur 2 à 5 et des patterns de longueurs 2 à 6 (avant et après correction)</i>	- 326 -	
<i>Annexe 10 Détail des patterns de chemins évalués (échantillon) pour l'analyse des résultats de l'extraction du réseau social et la détermination de la valeur seuil de filtrage du réseau social d'enquête.</i> -	328 -	
<i>Annexe 11 Détail des symboles minimaux dans l'échantillon et occurrences dans le fichier de données initial pré-traité</i>	- 330 -	

<i>Annexe 12</i>	<i>Scripts R écrits avec l'interface RStudio pour les générer les graphiques d'analyse du modèle...</i>	<i>331 -</i>
<i>Annexe 13</i>	<i>Interprétation des indicateurs de performance utilisés en complément des courbes ROC.</i>	<i>333 -</i>
<i>Annexe 14</i>	<i>Scripts R écrits avec l'interface RStudio pour effectuer les visualisations et calculer les centralités secondaires</i>	<i>334 -</i>
<i>Annexe 15</i>	<i>Centralité de degré des 62 acteurs du réseau pour lesquels cet indicateur prend des valeurs extrêmes (par rapport à l'ensemble des 615 acteurs)</i>	<i>342 -</i>
<i>Annexe 16</i>	<i>Centralité d'intermédiarité standardisée (pourcentage d'appartenance à l'un des chemins géodésiques existants entre deux nœuds quelconques) des 146 acteurs du réseau pour lesquels cet indicateur prend des valeurs extrêmes (par rapport à l'ensemble des 615 acteurs)</i>	<i>343 -</i>
<i>Annexe 17</i>	<i>Liste des acteurs des quatre profils de centralité d'intérêt</i>	<i>347 -</i>
Table des matières		350 -