



HAL
open science

Extraction d'un réseau social criminel par transformation d'un graphe d'enquête multivarié

Bruno Pinaud, Maud Bénichou, Guy Melançon

► **To cite this version:**

Bruno Pinaud, Maud Bénichou, Guy Melançon. Extraction d'un réseau social criminel par transformation d'un graphe d'enquête multivarié. Extraction et Gestion des Connaissances (EGC) 2023, Jan 2023, Lyon, France. pp.151-162. hal-03929950

HAL Id: hal-03929950

<https://hal.science/hal-03929950>

Submitted on 9 Jan 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.

Extraction d'un réseau social criminel par transformation d'un graphe d'enquête multivarié

Bruno Pinaud*, Maud Bénichou**,***, Guy Melançon*

* Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800
351 Cours de la Libération, 33405 Talence Cedex
{prenom.nom}@u-bordeaux.fr

** Univ. Bordeaux, IRM, Av Léon Duguit, CS 50057, 33608 Pessac
maud.benichou@ik.me

*** Service Central de Renseignement Criminel
Pôle Judiciaire de la Gendarmerie Nationale,
5 boulevard de l'Hautil 95001 Cergy Pontoise Cedex

Résumé. L'analyse des réseaux sociaux que constituent les organisations criminelles permet de capitaliser des connaissances sur leurs structures pour faciliter la détection de leurs acteurs essentiels. Cependant, les actes d'enquêtes établissent des liens de différentes natures (par ex. géolocalisation, propriété) entre différents types d'entités (par ex. personnes, lieux, véhicules). Ainsi se pose le défi d'extraire du graphe d'enquête multivarié un réseau social en considérant l'ensemble des informations à disposition. Il sera alors possible d'effectuer des analyses structurelles basées sur les mesures de centralité, appuyant l'identification des acteurs clés du réseau. Cet article propose une méthode d'extraction d'un réseau social à partir d'un tel graphe multivarié, les variables attachées aux sommets et arêtes du graphe multivarié étant prises en compte pour quantifier la vraisemblance des liens induits dans le réseau social.

1 Introduction

L'analyse des réseaux sociaux constitués par les organisations criminelles permet de capitaliser des connaissances sur leurs structures sociales, notamment au travers de la détection des acteurs essentiels à leur fonctionnement dans le but tant d'appuyer des enquêtes judiciaires que pour mettre en place des solutions proactives d'entrave des activités criminelles (Morselli, 2010; Everton, 2012). Cependant, les procès-verbaux d'actes d'enquêtes ne contiennent pas nécessairement de liens directs entre des personnes impliquées mais plutôt différents types de liens (géolocalisation, appels téléphoniques, *etc.*) entre différents types d'entités (personnes, lieux, véhicules, téléphones, *etc.*) (Lavaud-Legendre et al., 2017). Or, les indicateurs de centralité utiles pour caractériser les acteurs ou cerner leur position dans le réseau – telle la centralité d'intermédiarité (Freeman, 1977) — requièrent de disposer d'un réseau *social*, c'est-à-dire liant exclusivement des personnes entre elles (Bichler, 2019). Le calcul d'un tel réseau social n'est pas immédiat, sauf peut-être dans le cas où le réseau de départ est biparti, aussi appelé

2-modes (Borgatti, 2009), bien que ce cas qui semble plus facile ait donné lieu à différentes approches (Giatsidis et al., 2011; Neal, 2013).

Cet article propose une méthode de transformation d'un graphe multivarié G réunissant les différentes entités issues des actes d'enquêtes (voir un extrait figure 1a) en un réseau social N constitué par l'organisation criminelle (figure 1b). Pour chaque paire d'individus $\{u, v\}$ reliée par au moins un chemin dans G , la méthode consiste à induire un lien unique pondéré entre u et v dans N . La sémantique des éléments qui composent les différents chemins entre u et v dans G est prise en compte pour quantifier la vraisemblance des liens induits dans le réseau social N . Cette approche permet aux analystes d'avoir recours aux calculs usuels généralement utilisés sur les réseaux sociaux (Borgatti et al., 2018). La transformation est conduite de manière à historiser la création d'un lien dans N , permettant ainsi à un analyste d'examiner les éléments du graphe multivarié G ayant donné naissance aux liens dans le réseau social N .

Ainsi, les contributions de cet article :

- Apportent un cadre de lecture d'un réseau multivarié G facilitant sa transformation en un réseau social N entre individus, se prêtant dès lors aux méthodes d'analyse des réseaux sociaux (centralités, calcul de communautés, etc.).
- Explicitent un algorithme de transformation et son implémentation, avec le logiciel de visualisation TULIP (Auber et al., 2017)¹, applicable à tout graphe multivarié induisant à partir des chemins entre individus un lien direct unique entre eux.
- Illustrent et valident l'approche sur un jeu de données réelles anonymisées issues d'une enquête de gendarmerie et rendu disponible pour la communauté.

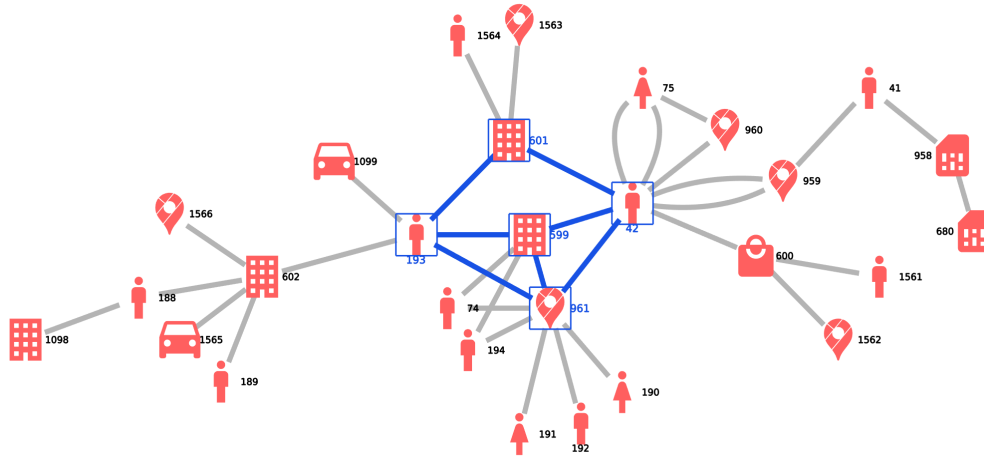
La section 2 présente un état de l'art des méthodes permettant d'induire un réseau social depuis un graphe multivarié et souligne la singularité de nos travaux. La section 3 présente les différentes phases de calcul de l'algorithme. La section 4 présente un cas d'application sur un réseau de terrain avant une conclusion en section 5.

2 État de l'art

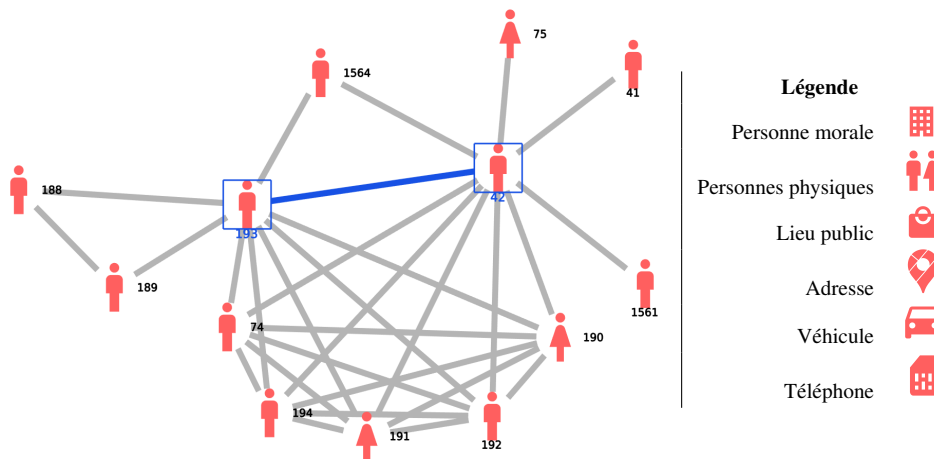
Nous n'avons pas connaissance d'algorithmes établis qui permettent d'induire à partir d'un graphe multivarié G un réseau social, c'est-à-dire de calculer à partir de $G = (V, E)$ un graphe $N = (P, F)$ rassemblant les sommets $P \subset V$ d'un type donné liés entre eux alors qu'ils ne le sont pas nécessairement dans G . Cela dit, la notion de contraction de chemins afin de réduire un graphe a fait l'objet de nombreux travaux et ce, depuis de nombreuses années (Aho et al., 1972; Martello, 1979), et demeure un sujet d'actualité (Bessouf et al., 2019).

Une situation usuelle proche est celle où le graphe multivarié G est un graphe biparti, aussi appelé 2-modes, ou réseau d'affiliation (Borgatti et al., 2018). Un tel graphe comporte deux types d'entités t et t' et les arêtes lient des entités qui sont nécessairement de types différents. C'est le cas, par exemple, lorsque des individus sont liés à des événements auxquels ils participent. Le réseau social N obtenu de G lie des personnes ayant co-participé à au moins un événement. Dans ce cas souvent rencontré, on dit parfois que N est obtenu par *projection* du graphe G . Les liens du réseau social peuvent être pondérés pour, notamment dans cet exemple, comptabiliser le nombre d'événements auxquels les individus ont co-participé.

1. <https://tulip.labri.fr>



(a) L'état initial du réseau multivarié G



(b) Le réseau social final N

FIG. 1 – (a) Extrait d'un graphe issu d'une enquête criminelle. Cet article décrit comment les éléments en bleus sont transformés pour créer un lien direct entre les personnes (b).

Dans le cas 2-modes, le passage au réseau social engendre un graphe d'une forte densité, essentiellement parce que le graphe résultant est en quelque sorte une "collection de cliques". En effet, pour reprendre notre exemple, les personnes ayant co-participé à un évènement sont alors toutes liées deux à deux. Pour pallier à cette situation, Neal (2013) propose un critère probabiliste qui permet de contrôler le nombre d'arêtes induites dans le réseau social. Giatsidis et al. (2011) proposent un calcul du poids des arêtes induites qui tient compte de la taille des cliques par lesquelles elles arrivent. Un seuil imposé aux arêtes permet alors de filtrer le graphe pour obtenir un réseau social d'une densité convenable. Dans une approche faisant une analogie avec la capacité à mémoriser une information à long terme, Kudelka et al. (2010)

étendent cette idée au cas où les données présentent un caractère temporel. On peut aussi citer Nick et al. (2013) qui proposent de filtrer un réseau social de forte densité en privilégiant la présence de triades pour en obtenir un "squelette".

Cet article se penche sur le cas de graphes multivariés où plus de deux types de sommets sont présents (notre cas d'application en compte 10, cf. section 4). De plus, ils ne présentent pas de régularités comme c'est le cas pour les graphes 2-modes, puisque les sommets d'un type donné peuvent être liés par des chemins de longueur variable. Dans le contexte ayant donné naissance à notre travail, la procédure calculant le poids d'une arête doit refléter combien il est vraisemblable que deux personnes soient effectivement liées l'une à l'autre (bien qu'aucun contact direct n'ait été observé), afin d'amener un analyste à lui accorder une certaine crédibilité dans son travail d'enquête. Le calcul du poids d'une arête dans le réseau social N est l'une des questions au cœur de l'algorithme que nous proposons.

3 Passer du multigraphe multivarié à un réseau social

Cette section pose quelques définitions avant de décrire les différentes transformations pour passer du graphe multivarié modélisant les relations établies durant l'enquête au réseau social.

3.1 Graphe multivarié, force liante et sommets projetés

Le graphe d'enquête est modélisé à l'aide d'un multigraphe $G = (V, E)$ non orienté et sans boucle avec V l'ensemble des sommets et E l'ensemble des arêtes (aussi appelées *liens*). Dans un multigraphe, deux sommets $\{u, v\} \in V$ peuvent être liés par plus d'une arête. Cela se justifie dans le contexte d'une enquête, par exemple, lorsqu'un individu est observé à de multiples occasions en un même lieu donnant ainsi naissance à des *arêtes multiples* liant cet acteur à ce lieu.

Une sémantique est associée aux éléments de V et E , faisant du multigraphe G un *réseau* au sens de Wasserman et Faust (1994). Plus précisément, un sommet $u \in V$ (resp. une arête $e \in E$) a un *type* t_u (resp. t_e). Ainsi, dans le cas d'usage qui sera présenté à la section 4, la catégorie *PERSONNE* correspond à l'ensemble des sommets $u \in V$ pour lesquels $t_u = PERSONNE$. Il en va de même pour les autres types, qui définissent des catégories *LIEU*, *VEHICULE*, etc. De la même manière, on peut considérer des catégories de liens comme *UTILISATEUR* (d'un véhicule, d'un téléphone), *FAMILIAL* (deux personnes de la même famille), *LOCALISATION* (pour une personne vue en un lieu donné), etc.

Étant donné un lien $e = (u, v)$, on définit la *force liante* $f(t_u, t_e, t_v) \in [0, 1]$ associée au triplet que forment e et ses deux extrémités $\{u, v\}$ (par souci de simplicité dans la suite, les triplets sont désignés uniquement par les liens dont ils sont composés). La force liante, définie par l'analyste, reflète la capacité du triplet, en tant que fragment d'un chemin liant deux personnes, à soutenir l'existence d'un lien direct entre elles. Par exemple, pour l'analyste, le fait que deux individus aient été localisés au même moment dans un lieu privé (une maison), soutient fortement l'existence d'un lien direct entre elles et incite à attribuer une valeur élevée à la force liante $f(PERSONNE, LOCALISATION, LIEU PRIVE)$ (0.7 dans le cas d'usage présenté à la section 4). En revanche, le fait que deux véhicules utilisés par deux individus aient été localisés dans un lieu public (parking d'un centre commercial) sou-

tient modérément l'existence d'un lien direct, suggérant d'attribuer une valeur plus faible à f (*VEHICULE, LOCALISATION, LIEU PUBLIC*) (0.3 dans le cas d'usage présenté à la section 4).

Les sommets qui deviendront les sommets du réseau social sont appelés les *sommets projetés* (ceux qui font l'objet de la projection sur le réseau social N), et nous désignons par P l'ensemble de ces sommets. Traditionnellement, les sommets projetés représentent des personnes. Néanmoins, dans le cas d'un réseau de trafics, il pourrait être intéressant de rapprocher des objets (voitures, substances illicites, etc.) afin d'être éclairé sur le mode opératoire des trafiquants ou de détecter des séries.

Les sommets de $V \setminus P$ qui sont d'un autre type seront appelés des *sommets accessoires*. Nous nous intéressons aux chemins liant des sommets $u, v \in P$, et qui ne passent que par des sommets accessoires (hormis leurs extrémités u, v). En d'autres mots, seules les extrémités de ces chemins sont des sommets projetés ; nous dirons que ces chemins sont *admissibles*.

Ce faisant, le réseau est construit sans égard aux triades (ou plus précisément aux chemins liant simultanément trois sommets projetés à la fois). Dans un article fondateur, Granovetter (1973) met en exergue une configuration qualifiée d'impossible dans un réseau social : si un individu X possède un lien fort avec des individus Y et Z , alors il n'est pas possible que Y et Z ne soient pas liés. Cependant, dans un réseau criminel, la posture des protagonistes est différente. Afin de conserver à la fois leur importance stratégique et leurs éventuelles commissions, ils n'ont aucun avantage à faire bénéficier les autres membres de l'organisation de leurs contacts. De plus, réduire le nombre d'individus ayant accès aux autres membres, plans ou opérations aide à protéger le réseau dans son ensemble (DoD, 2016). Ainsi si Y est lié à X lui-même lié à Z , cela ne permet pas de supposer l'existence d'un lien direct entre Y et Z .

3.2 Fusion de liens, élagage et contraction de chemins

Classiquement en analyse de réseaux sociaux, un analyste recherche le ou les plus courts chemins entre deux individus (Wasserman et Faust, 1994), privilégiant parfois les chemins avec la meilleure pondération. *A contrario*, notre objectif, qui vise à extraire un réseau social depuis un graphe multivarié, ne se limite pas aux seuls plus courts chemins dans le multigraphe. Cela permet de prendre en compte toutes les dimensions "sémantiques" qui entrent en jeu dans la relation entre les sommets projetés. Dès lors que deux sommets projetés $u, v \in P$ sont liés par un chemin dans le multigraphe G , ils seront liés dans le réseau social N (Figure 1). La force liante de l'arête liant u et v dans N résulte d'opérations appliquées sur les forces liantes des différentes arêtes de G lui donnant naissance.

Une première phase de *fusion* traite le cas des arêtes multiples. On convient qu'une arête multiple entre deux sommets de $u, v \in V$ correspond à un multi-ensemble d'arêtes que nous noterons simplement $\{e_1, \dots, e_k\}$. Une première phase remplace chaque arête multiple par une arête "simple" $e = (u, v)$ dont la force est alors définie comme étant $f = \phi(f_1, \dots, f_k)$, où ϕ est un opérateur associatif et commutatif, à l'image de l'opération arithmétique d'addition. Cela permet d'étendre sans ambiguïté l'opérateur ϕ à un nombre indéterminé d'opérandes. La figure 1a comporte deux cas entre les sommets 42 (entouré en bleu), 75 et 959. Dans la figure 2a, les arêtes simples fusionnées sont de couleur bleue.

Après cette première phase, le graphe est simple (ni arête multiple, ni boucle). Il est susceptible de contenir des sommets accessoires de degré 1. Ces sommets sont un artefact des données d'observation et ne contribuent pas à renforcer des liens entre des sommets projetés. En accord avec la logique qui cherche à créer le réseau social entre des sommets projetés, le

Extraction d'un réseau social criminel

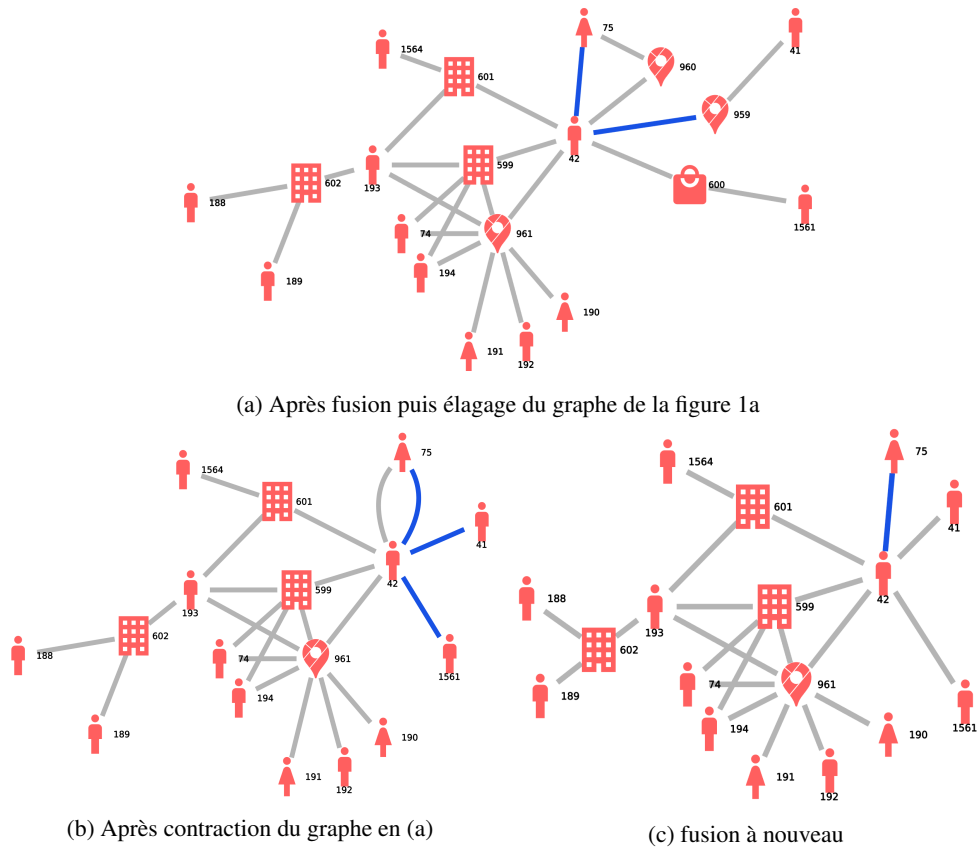


FIG. 2 – Évolution du graphe de la figure 1a issu du cas d'application détaillé dans la section 4. Les arêtes bleues sont issues de la transformation illustrée par rapport à l'étape précédente.

choix est fait de les supprimer. Ce faisant, la suppression d'un sommet accessoire de degré 1 peut diminuer le degré de son voisin, accessoire aussi et en faire un sommet de degré 1 à son tour. On le supprime aussi, et ainsi de suite jusqu'à ne plus en avoir parmi les sommets accessoires. Cette phase, effectuée entre les figures 1a et 2a (par ex. les sommets 958 et 680 à droite de la figure 1a), est appelée l'*élagage*, puisque la succession de sommets ainsi supprimés forme une chaîne apparaissant comme une "branche" accrochée au reste du graphe.

Après élagage, un chemin admissible peut contenir des sommets accessoires de degré 2 (par ex. le sommet 959 de la figure 2a). Soit v un tel sommet lié à deux autres sommets u, w , **et à eux seuls**, par des arêtes $e = (u, v)$ et $e' = (v, w)$ de forces liantes respectives f, f' . Les deux arêtes e, e' sont alors *contractées* pour donner lieu à une arête $e'' = (u, w)$ dont la force liante est alors égale à $f'' = f \cdot f'$ (figure 2b). Cette étape peut créer des arêtes multiples entre les sommets u et w lorsque le sommet v de degré 2, supprimé par l'effet de la contraction, forme une triade avec les sommets u et w (entre les sommets 42 et 75 sur la figure 2b). La contraction doit donc être suivie d'une fusion afin de se trouver à nouveau avec

un graphe simple, avant de poursuivre l'étape de contraction (figure 2c). A noter toutefois que la propriété d'associativité de l'opérateur ϕ permet de retarder l'étape de fusion, à condition d'en tenir compte lorsqu'on évaluera le degré des sommets u ou w qui pourront faire l'objet ultérieurement d'une contraction.

Il y a donc nécessité d'itérer les étapes de fusion, d'élagage et de contraction des sommets de degré 2 jusqu'à ce le graphe résultant ne contienne aucune occurrence d'arêtes multiples à fusionner, de sommets accessoires de degré 1 à supprimer ou de sommets accessoires de degré 2 donnant lieu à une contraction d'arêtes.

3.3 Contraction de sous-graphes

On considère ensuite, pour chaque paire de sommets projetés $u, v \in P$ le sous-graphe $H_{u,v}$ de G composé de l'union de l'ensemble des chemins admissibles et élémentaires entre u et v (qui ne passent pas deux fois par le même sommet). La figure 3 montre 3 exemples issus de la figure 2c où les sommets u, v sont de type *PERSONNE*. Parce que les sommets font maintenant partie d'un sous-graphe de G , il est possible que des sommets accessoires soient de degré 2 dans $H_{u,v}$ (sommets 599 de la figure 3b à comparer avec la figure 2c). Il faut donc itérer sur $H_{u,v}$ les opérations de contraction et de fusion. Chaque sous-graphe $H_{u,v}$ pourra alors être contracté jusqu'à n'avoir plus qu'une seule arête entre les sommets projetés $u, v \in P$. Cette arête devient dès lors une arête du réseau social, dont la force liante aura été calculée au fil de la projection comme précisé à la section 3.2.

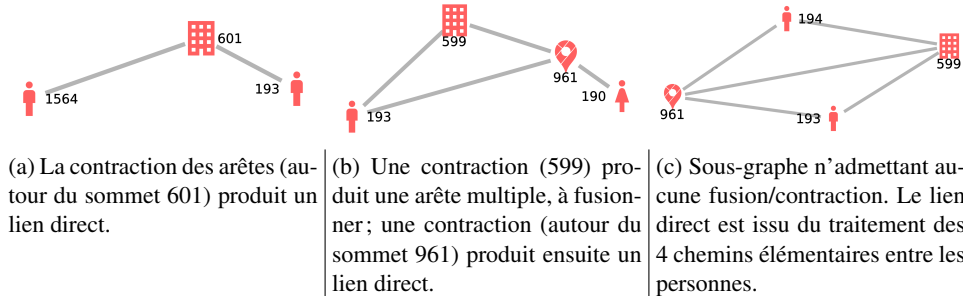


FIG. 3 – Quelques sous-graphes $H_{u,v}$ issus de l'exemple de la figure 2c.

Cependant, certains sous-graphes ne se réduisent pas à une arête entre les sommets projetés $u, v \in P$, comme illustré à la figure 3c. Pour ceux-ci, on énumère alors l'ensemble des chemins $\{c_1, c_2, \dots, c_k\}$ élémentaires entre u et v et on les contracte pour calculer pour chacun sa force liante f_1, f_2, \dots, f_k , égale au produit des forces liantes des arêtes qui le forment. On retire le chemin c_{\max} de force liante f_{\max} maximale parmi tous les chemins considérés. Les arêtes du chemin c_{\max} sont ensuite retirées du sous-graphe $H_{u,v}$ (on retire aussi les sommets qui ce faisant ne sont incidents à aucune arête) pour obtenir un sous-graphe $H'_{u,v}$. Cette phase de *réduction* est répétée jusqu'à épuisement des chemins élémentaires entre u et v , pour obtenir une séquence de chemin $c_{\max}, c'_{\max}, c''_{\max}, \dots$ extraits de sous-graphes $H_{u,v}, H'_{u,v}, H''_{u,v}, \dots$ de force liante $f_{\max}, f'_{\max}, f''_{\max}, \dots$. En définitive, une arête entre les sommets $u, v \in P$ est créée dans le réseau social N de force liante $\phi(f_{\max}, f'_{\max}, f''_{\max}, \dots)$.

3.4 Confluence des calculs

L'algorithme effectue ainsi une série d'étapes de fusions et d'élagages simplifiant en quelque sorte le graphe multivarié avant d'en contracter les arêtes par itérations successives. Cette étape est d'une complexité de $O(|E|)$ si on tient compte des arêtes multiples qui peuvent éventuellement faire l'objet d'une fusion, ou des paires d'arêtes qui sont contractées (suivies d'une fusion). La dernière étape, qui procède par énumération des chemins élémentaires dans les sous-graphes $H_{u,v}$, reste d'une complexité maîtrisée dès lors que ces sous-graphes sont de petite taille, ce vers quoi amènent les étapes de réduction qui la précèdent.

La question de la confluence mérite qu'on s'y arrête car l'ordre de traitement des arêtes du multigraphe lors des étapes de fusion, d'élagage et de contraction pourrait *a priori* influencer sur le résultat final. Or, ici entrent en jeu les propriétés de commutativité et d'associativité de l'opérateur ϕ . Pour s'en convaincre, il suffit d'examiner la situation où deux sommets projetés u, v sont liés par trois chemins distincts passant par des sommets accessoires w_1, w_2 et w_3 (figure 4). En effet, sans égard à l'ordre dans lequel les contractions, suivies de fusion, sont appliquées, la valeur associée à l'arête liant u et v dans le réseau social sera égale à $\phi(f_i \cdot f'_i, \phi(f_j \cdot f'_j, f_k \cdot f'_k))$ qui est aussi égale à $\phi(\phi(f_i \cdot f'_i, f_j \cdot f'_j), f_k \cdot f'_k)$, et ce pour tout choix de valeurs $\{i, j, k\} = \{1, 2, 3\}$ précisant l'ordre de traitement des arêtes à contracter. L'argument vaut aussi pour le cas où les contractions doivent s'opérer sur des chemins plus longs dont les valeurs associées sont le produit des forces liantes des arêtes qui les constituent.

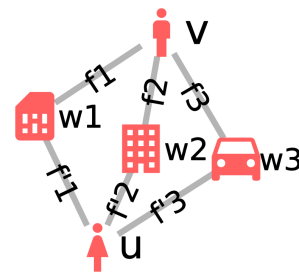


FIG. 4 – L'ordre dans lequel les trois chemins sont contractés n'importe pas en vertu des propriétés de l'opérateur ϕ .

3.5 Historisation des calculs

La définition de l'opérateur ϕ utilisé dans le calcul de la force liante d'une arête du réseau social est laissée à l'analyste, et dépend du domaine étudié – seules sont imposées les propriétés arithmétiques garantissant la confluence des calculs et l'appartenance des forces liantes à $[0, 1]$. Cela dit, les forces liantes ainsi calculées n'ont pas valeur de vérité et doivent pouvoir être examinées de près par l'analyste qui peut en questionner la vraisemblance.

À cette fin, il est possible de stocker au fil du déroulement de l'algorithme les informations nécessaires pour exhiber à l'analyste l'ensemble des chemins qui sont entrés en jeu dans le calcul de la force liante d'une arête du réseau social. Le logiciel TULIP (Auber et al., 2017) offre un modèle de données particulièrement adapté à cette fin. TULIP permet de gérer une hiérarchie de sous-graphes et des collections de propriétés associées aux éléments du graphe à la manière du mécanisme d'héritage des langages de programmation à objets. Ainsi, un sommet présent dans un sous-graphe n'est pas dupliqué en mémoire puisque d'un point de vue informatique il s'agit du même objet que dans son graphe parent. Un sous-graphe est créé en appliquant un filtre booléen sur son parent et les propriétés associées aux éléments du sous-graphe sont héritées du graphe parent.

TULIP permet de présenter simultanément deux vues, l'une sur le réseau social, l'autre sur le graphe multivarié initial. Le mécanisme d'héritage autorise alors, par une sélection simple

dans le réseau social de sommets projetés u, v , d’identifier ces mêmes acteurs dans le graphe multivarié. Les informations stockées sur la cascade d’opérations de fusion, élagage et contraction doublées de l’accès au sous-graphe $H_{u,v}$ rend possible la production à la volée du sous-graphe ayant donné lieu à l’arête (u, v) , qui peut alors être examiné par l’analyste.

4 Cas d’usage sur des données d’enquêtes réelles

Nous rapportons ici des résultats de l’algorithme appliqué sur un multigraphe multivarié construit à partir de données d’enquête. L’experte en sciences criminelles, co-auteure de l’article, auprès de laquelle les données ont été obtenues a de plus participé activement à la conception de l’algorithme de la section 3. Les données proviennent d’une enquête de gendarmerie sur un réseau de trafic de fonds et de blanchiment ayant agi sur l’ensemble du territoire français. Une version anonymisée des données est librement accessible (Pinaud et al., 2022).

Le graphe multivarié initial G possède 1703 sommets de 10 types différents dont 665 de type *PERSONNE* qui forment l’ensemble des sommets projetés P . Il compte 2318 arêtes pour 19 types de liens. Chaque arête est associée à une force liante définie par les analystes. Le réseau possède 12 composantes connexes pour un diamètre de 23 et une longueur moyenne des chemins proche de 4. Le degré moyen des sommets est 2.7 (écart-type 4.9). La faible densité d’arêtes par rapport au nombre de sommets est une caractéristique commune des réseaux criminels (Morselli, 2009). Dans notre cas, G ne contient que les liens présentant un intérêt pour les forces de l’ordre et validés dans le cadre de l’enquête judiciaire.

La fusion des liens utilise une opération ϕ définie par les experts qui reflète leur importance relative. Soit la force liante f d’un lien de G , on définit une première fonction :

$$\gamma(f) = \begin{cases} 0.01 & \text{si } 0 < f < \alpha \\ 0.05 & \text{si } \alpha \leq f < \beta \\ 0.1 & \text{si } \beta \leq f < 1 \end{cases}$$

où α, β sont définies à la discrétion de l’expert, de manière à classer les arêtes du graphe en liens faiblement ($0 < f < \alpha$), moyennement ($\alpha \leq f < \beta$) ou fortement ($\beta \leq f < 1$) liants. On pose ensuite $\varphi(f, f') = \gamma(f) + \gamma(f')$. On vérifie facilement que l’opérateur φ est commutatif et associatif. On observe que les fonctions max et min vues comme des opérateurs sont elles aussi commutatives et associatives. Il en va donc de même de l’opérateur $\phi(f, f') = \min(\max(f, f'), \varphi(f, f'))$ défini par nos experts pour le graphe G de ce cas d’usage. Au dire de notre experte en science criminelle, cette possibilité de façonner l’opérateur ϕ contribue en quelque sorte à rapprocher le calcul de son interprétation.

Après la première phase de *fusion-élagage-contraction* décrite dans la section 3.2, les nombres de sommets et d’arêtes ont diminué de près de moitié (832 sommets, 1136 arêtes) et 4 composantes connexes contiennent encore des sommets accessoires. Le diamètre du graphe passe à 17, la longueur moyenne des chemins à 3.2, et le degré moyen des sommets à 2.5 (écart-type 2.4). L’étape suivante, qui construit les sous-graphes $H_{u,v}$ (et extrait de chacun l’ensemble des chemins élémentaires), génère 12996 sous-graphes – qui résulteront chacun en une arête du réseau social N – dont le plus gros a 37 sommets et 58 arêtes. Parmi ceux-ci, 8429 sous-graphes se réduisent à un lien direct entre deux personnes. La réduction des 4567 sous-graphes restant irréductibles par fusion/contraction nécessite d’énumérer, réduire et fu-

sionner 39258 chemins. Le plus grand de ces sous-graphes $H_{u,v}$ possède encore 14 sommets et 24 arêtes.

En définitive, le réseau social N possède 665 sommets et 12996 arêtes. Le degré moyen des sommets est donc de 40 (chaque personne est en lien avec 40 autres personnes). Ce fort degré est à l'image de la projection d'un graphe 2-modes en un graphe 1-mode, faisant passer un graphe en étoile comptant N arêtes à une clique de taille $O(N^2)$. Dans notre cas, l'explosion combinatoire reste maîtrisée puisque le nombre d'arêtes est multiplié par 6. Cela dit, l'analyse du réseau social, notamment avec les techniques classiques (Morselli, 2009), requiert d'en filtrer les arêtes. Cette opération, laissée à la discrétion des experts, pourra s'appuyer sur une inspection des liens en ayant recours à l'historique de la projection (cf. section 3.5).

La figure 1 illustre notre contribution qui permet à l'analyste de solliciter les mesures de centralités usuelles appliquées à un réseau social, sans pour autant "oublier" les observations de terrain desquelles sont induites les liens. En effet, équipé d'une vue multiple et grâce au mécanisme d'historisation, il devient possible pour l'analyste de calculer et visualiser le sous-graphe de G ayant donné naissance à une arête du réseau social N .

L'algorithme a été implémenté à l'aide de la librairie `tulip-python` (Lambert et Auber (2012)), librairie compagne du logiciel TULIP). Il a été exécuté sur un seul cœur d'un ordinateur récent (i7-11850H à 2.50GHz, 32Go RAM, TULIP 5.7, Python 3.10). L'étape de fusion-élagage-contraction prend moins d'une seconde. Le calcul des sous-graphes $H_{u,v}$, et plus encore l'énumération pour chacun des chemins élémentaires requiert un peu moins de 6 minutes, dont moins d'une minute pour traiter le cas des 8429 sous-graphes réduits à un lien direct entre deux personnes. Ce temps de calcul reste raisonnable pour l'utilisateur puisque l'algorithme vient en étape préalable à l'analyse. On le doit à l'évidence à la faible densité du multigraphe multivarié G .

5 Conclusion

Nous avons présenté un algorithme qui à partir d'un multigraphe multivarié modélisant les relations établies au cours d'une enquête criminelle permet de calculer un réseau social entre les personnes qui se trouvent connectées par des chemins liant différents éléments de l'enquête (lieux, véhicules, etc.). Le déroulement de l'algorithme permet d'attribuer aux liens du réseau social une force liante, qui quantifie leur vraisemblance sur le terrain. La définition de la force liante des arêtes du multigraphe de départ appartient à l'expert de domaine. Un opérateur, aussi défini par l'expert, permet de faire la synthèse des forces liantes des chemins induisant une force liante sur les liens du réseau social. Cette définition doit répondre à certaines conditions algébriques – en tous points similaires à l'opération arithmétique d'addition, pour assurer la confluence des calculs. L'historisation des calculs (rendue possible par les capacités de la librairie TULIP à gérer une hiérarchie de sous-graphes) apporte à l'algorithme sa capacité à expliquer les choix qui ont été faits pour induire un lien du réseau social.

L'algorithme, présenté dans le contexte de l'analyse criminelle et de la construction d'un réseau social impliquant des personnes, est générique. Il peut calculer un réseau liant directement des entités d'un même type, quel qu'il soit, plongées au départ dans un multigraphe multivarié. Ainsi, afin de questionner les modes opératoires d'un réseau criminel, on pourrait calculer un réseau "social" rapprochant des substances illicites apparaissant au départ comme

des entités d'un graphe multivarié qui rend compte d'observations récoltées au cours d'une enquête de cyber-criminalité.

Un développement futur de notre approche est la prise en compte de la dimension temporelle des liens (par ex. date/heure de présence en un lieu), ce à quoi ne se prêtaient pas les données auxquelles nous avons accès. Nous envisageons aussi d'appliquer l'algorithme à l'analyse de cyber-criminalité (Rossy et Décary-Héту, 2017; Rhumorbarbe et al., 2018) et plus généralement dans les humanités numériques où la construction d'un réseau social à partir de données multivariées reste un problème complexe (Pister et al., 2022).

Références

- Aho, A. V., M. R. Garey, et J. D. Ullman (1972). The transitive reduction of a directed graph. *SIAM Journal on Computing* 1(2), 131–137.
- Auber, D., D. Archambault, R. Bourqui, M. Delest, J. Dubois, A. Lambert, P. Mary, M. Mathiaut, G. Mélançon, B. Pinaud, B. Renoust, et J. Vallet (2017). TULIP 5. In R. Alhajj et J. Rokne (Eds.), *Encyclopedia of Social Network Analysis and Mining*, pp. 1–28. Springer.
- Bessouf, O., A. Khelladi, et T. Zaslavsky (2019). Transitive closure and transitive reduction in bidirected graphs. *Czechoslovak Mathematical Journal* 69(2), 295–315.
- Bichler, G. (2019). *Understanding Criminal Networks : A Research Guide*. Univ. of California Press.
- Borgatti, S. P. (2009). 2-mode concepts in social network analysis. *Encyclopedia of complexity and system science* 6, 8279–8291.
- Borgatti, S. P., M. G. Everett, et J. C. Johnson (2018). *Analyzing Social Networks, 2nd edition*. Sage.
- DoD, U. S. (2016). *Countering Threat Networks*. CreateSpace Independent Publishing Platform. https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_25.pdf.
- Everton, S. F. (2012). *Disrupting dark networks*. Cambridge Univ. Press.
- Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry* 40(1), 35–41.
- Giatsidis, C., D. M. Thilikos, et M. Vazirgiannis (2011). Evaluating cooperation in communities with the k -Core structure. In *Int. Conf. on advances in social networks analysis and mining*, pp. 87–93. IEEE.
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology* 78(6), 1360–1380.
- Kudelka, M., Z. Horak, V. Snasel, et A. Abraham (2010). Social network reduction based on stability. In *Int. Conf. on Computational Aspects of Social Networks*, pp. 509–514. IEEE.
- Lambert, A. et D. Auber (2012). Graph analysis and visualization with Tulip-Python. In *EuroSciPy 2012 - 5th European meeting on Python in Science*.
- Lavaud-Legendre, B., C. Plessard, A. Laumond, G. Melançon, et B. Pinaud (2017). Analyse de réseaux criminels de traite des êtres humains : méthodologie, modélisation et visualisation. *J. of Interdisciplinary Methodologies and Issues in Science* 2. *Graphes et systèmes sociaux*.

- Martello, S. (1979). An algorithm for finding a minimal equivalent graph of a strongly connected digraph. *Computing* 21(3), 183–194.
- Morselli, C. (2009). *Inside criminal networks*, Volume 8 of *Studies of Organized Crime*. Springer New York.
- Morselli, C. (2010). Assessing vulnerable and strategic positions in a criminal network. *Journal of Contemporary Criminal Justice* 26(4), 382–392.
- Neal, Z. (2013). Identifying statistically significant edges in one-mode projections. *Social Network Analysis and Mining* 3, 915–924.
- Nick, B., C. Lee, P. Cunningham, et U. Brandes (2013). Simmelian backbones : Amplifying hidden homophily in facebook networks. In *Advances in Social Networks Analysis and Mining (ASONAM), IEEE/ACM Int. Conf. on*, pp. 525–532.
- Pinaud, B., M. Bénichou, et G. Melançon (2022). Multivariate criminal investigation network [data set]. <https://doi.org/10.5281/zenodo.7044351>.
- Pister, A., N. Dufournaud, P. Cristofoli, C. Prieur, et J.-D. Fekete (2022). From Historical Documents To Social Network Visualization : Potential Pitfalls and Network Modeling. In *VIS4DH 2022 - 7th Workshop on Visualization for the Digital Humanities*, Oklahoma City, United States.
- Rhumorbarbe, D., D. Werner, Q. Gilliéron, L. Staehli, J. Broséus, et Q. Rossy (2018). Characterising the online weapons trafficking on cryptomarkets. *Forensic science international* 283, 16–20.
- Rossy, Q. et D. Décary-Héту (2017). Internet traces and the analysis of online illicit markets. In *The Routledge international handbook of forensic intelligence and criminology*, pp. 249–263. Routledge.
- Wasserman, S. et K. Faust (1994). *Social Network Analysis : Methods and Applications*. Structural Analysis in the Social Sciences. Cambridge Univ. Press.

Summary

Analysis of social networks constituted by criminal organizations can capitalize on knowledge about their structures to facilitate the detection of their key actors. However, investigative acts establish links of different natures (e.g., geolocation, ownership) between different types of entities (e.g., people, places, vehicles). Thus arises the challenge of extracting the social network from the multivariate investigation graph by considering all available information. It is then possible to perform structural analyses based on centrality measures, supporting the identification of key actors in the network. This paper proposes a method for extracting a social network from such a multivariate graph, the variables attached to the nodes and edges of the multivariate graph being taken into account to quantify the likelihood of the links induced in the social network.