HAL open science

Pratiques de l'analyse de réseau en géographie. Retours sur trois thèses récentes

Élise Autrive, Laurent Beauguitte, Ninon Briot, Paul Gourdon

To cite this version:

Élise Autrive, Laurent Beauguitte, Ninon Briot, Paul Gourdon. Pratiques de l'analyse de réseau en géographie. Retours sur trois thèses récentes. 2023. hal-03920666v1

HAL Id: hal-03920666

https://hal.science/hal-03920666v1

Preprint submitted on 3 Jan 2023 (v1), last revised 27 Feb 2024 (v4)

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.

Pratiques de l'analyse de réseau en géographie
Retours sur trois thèses récentes

Élise Autrive*, Laurent Beauguitte**, Ninon Briot*** et Paul Gourdon**

* UMR IDEES **UMR Géographie-cités ***ENS Lyon

Proposition d'article pour le dossier « Concepts et méthodes en analyse de réseaux : circulations et barrières disciplinaires » de la revue en ligne *ARCS.*

Résumé : Trois géographes ayant soutenu en 2021 et mobilisant des méthodes d'analyse de réseau issues d'autres disciplines (sociologie, informatique, physique) expliquent le choix des méthodes, des indicateurs et des logiciels utilisés dans leurs travaux. Iels insistent à la fois sur la faiblesse de la formation initiale et le rôle crucial de l'auto-formation. Les forces et faiblesses de l'analyse de réseau et des visualisations de type liens-nœuds sont enfin évoqués.

Mots clés : Géographie, Analyse de réseau, (Auto) Formation, Logiciels, Visualisation

Abstract : Three geographers who have defended their thesis in 2021 and who use network analysis methods from other disciplines (sociology, computer science, physics) explain the choice of methods, indicators and software used in their work. They insist on both the weakness of initial training and the crucial role of self-training. The strengths and weaknesses of network analysis and link-node visualizations are finally discussed.

Key words: Geography, Network analysis, (Self-) Training, Softwares, Visualization

L'analyse de réseau en géographie a une longue histoire. La tradition disciplinaire construite dans les années 1960 par des géographes quantitativistes nord-américains s'est concentrée sur deux types d'objets principaux : les réseaux d'infrastructures pouvant être modélisés sous forme de graphes planaires ; la simplification et/ou la modélisation des flux entre espaces (souvent des villes). Depuis plus d'une décennie, les hybridations disciplinaires liées à l'analyse de réseau se font plus nombreuses dans les thèses de géographie soutenues en France. L'utilisation de méthodes issues de la physique ou de l'analyse des réseaux sociaux, si elle n'est pas tout à fait banalisée, est devenue semble-t-il plus fréquente ces dernières années. Trois thèses soutenues en 2021 en géographie montrent cette hybridation en actes : les mesures sont issues de l'analyse de réseaux sociaux et/ou de la physique des réseaux ; les concepts développés dans d'autres disciplines sont retravaillés en fonction des

problématiques spatiales des auteurs et autrices ; les enjeux liés à l'analyse et à la visualisation des données relationnelles sont explicitement évoqués.

Ces trois thèses, disponibles en ligne, sont issues de trois laboratoires différents et les démarches ne sont pas totalement comparables. Élise Autrive (UMR IDEES) s'intéresse à l'accès aux soins en Guyane et mobilise l'analyse de réseaux sociaux pour d'une part connaître la structure des relations entre acteurs du système sanitaire et social et d'autre part les recours préférentiels à ces acteurs de la part de populations précaires. Ses données relationnelles sont issues d'enquêtes de terrain et la taille des réseaux étudiés reste modeste. Ninon Briot (ENS Lyon) et Paul Gourdon (UMR Géographie-cités) travaillent sur les relations entre villes européennes en étudiant leur participation à des projets de coopération : les données sont beaucoup plus volumineuses et issues de sources institutionnelles. Au-delà des différences liées aux questions de recherche et aux parcours personnels et académiques de ces trois géographes, un certain nombre de similarités apparaissent, qu'il s'agisse de la formation - et de l'importance de l'auto-formation -, des mesures et outils utilisés, des références mobilisées ou des difficultés rencontrées.

Les figures extraites de leurs thèses et qui ponctuent l'entretien permettent d'expliciter la démarche choisie ou de mettre en évidence les modélisations graphiques élaborées par l'auteur et les autrices. Les […] signalent des coupures dans les réponses fournies. Une série de questions a été envoyée par mail en juillet 2022 ; une fois les réponses collectées, des demandes de clarification ou des suggestions de reformulation ont été adressées à chacun.e. La mise en forme finale a été mise à disposition des auteurs et autrices afin qu'iels puissent, si iels le souhaitaient, développer – ou raccourcir – telle ou telle réponse. La version déposée a été validée par l'ensemble des auteurs et autrices.

Laurent Beauguitte (LB) *: L'analyse de réseau occupe une place importante dans votre thèse, est-ce quelque chose que vous aviez prévu dès le début de celle-ci ? Si oui, pourquoi ? Si non, vous rappelez-vous quand vous avez choisi cette démarche ?*

Élise Autrive (EA) : L'analyse de réseau n'était pas nécessairement prévue lorsque le projet de thèse a été écrit. Entre le projet et le terrain, des ajustements ont eu lieu très tôt, au commencement de la thèse ; ces ajustements ont également concerné la méthodologie du recueil des données.

J'ai choisi de réaliser des entretiens auprès des acteurs investis dans la structuration de la politique régionale de santé (Agence Régionale de Santé) et auprès des acteurs de soins de santé (institutionnels publics et associatifs) qui agissent auprès des populations en situation de vulnérabilité (personnes migrantes). L'objectif central de ces entretiens était d'identifier les collaborations établies entre acteurs pour la prise en charge de ces populations. L'analyse de réseau paraissait une méthode adéquate

pour les représenter, les visualiser et aller plus loin *via* les mesures et les méthodes d'analyse disponibles. Par ailleurs, au fil de la thèse, une enquête auprès des personnes en situation de vulnérabilité (personnes migrantes dans les quartiers de bâti spontané de la Communauté d'Agglomération Centre Littoral de Cayenne) a été réalisable. Là encore, j'ai décidé d'utiliser l'analyse de réseau pour identifier les lieux de soins préférentiels des populations pour différents types de soins de santé. Ces différents matériaux et analyses ont pu se compléter les uns et les autres pour éclairer la structuration des différents graphes.

Ninon Briot (NB) : L'analyse de réseau était prévue dès mon projet de thèse car le sujet s'y prêtait tout particulièrement. Effectivement, l'idée de départ est bien de chercher à visualiser des liens politiques entre villes qui n'avaient jamais été étudiés sous cet angle. Aussi, l'analyse de réseau nous semblait prometteuse car elle permettait de représenter ces liens, en alternant entre des analyses propres à la structure de cet ensemble de relations, mais aussi des analyses spatiales. Pour autant, mon niveau en analyse de réseau avant ma thèse était très faible et je n'avais donc qu'une vague idée des traitements possibles. Je n'avais en effet qu'une image très large des mesures existantes mais aussi de l'agencement des données nécessaires pour l'analyse de réseau. De ce fait, entre mon projet de thèse et mon manuscrit final, les ambitions mises dans les analyses de réseaux pour la compréhension de mon objet d'étude étaient très différentes.

Paul Gourdon (PG) : Ça remonte à loin maintenant, il est donc possible que ma réponse soit en grande partie une construction rétrospective. La réponse est oui car, dès mon projet de thèse pour le M2 Géoprisme de Paris 1, je mentionne explicitement l'analyse de réseau (AR) dans la partie méthodologie. Je pense que l'on peut dire qu'il y a deux grandes raisons à cela. D'abord, mon sujet ayant trait à la coopération transnationale entre villes européennes, je suis rapidement tombé sur le vocable un peu flou de « réseaux de villes » (et en anglais *Transnational Municipal Networks – TMN*). Lorsqu'on y regarde de plus près, ces termes renvoient simplement à des associations de municipalités mais la popularité du mot « réseau » et cette expression de « réseaux de villes » expliquent en partie pourquoi, par une sorte de glissement (et sans doute au départ pour des effets d'affichage, c'est-à-dire pour obtenir un contrat doctoral), j'ai mentionné l'AR, sans savoir en détail ce que cela recouvrait.

La deuxième raison provient là encore de la bibliographie : ces fameux « réseaux de villes » étaient souvent mobilisés dans des articles de science politique ou d'histoire comme une modalité d'européanisation parmi d'autres. Pour autant, dans la plupart des ouvrages de synthèse, dont *Le retour des villes européennes* de Patrick Le Galès (2011) est un bon exemple, on a presque toujours affaire aux mêmes associations et celles-ci sont simplement présentées sous forme de listes, afin de montrer leur caractère pléthorique. Cette relative frustration m'a conduit à vouloir « voir » à quoi ressemblerait le réseau d'un ensemble vaste d'associations de villes, sur le terrain européen. Dans mon projet de thèse, je parlais de réseaux « inter-organisationnels »

pour évoquer les types de liens que je pouvais reconstruire en regardant les villes membres d'un ensemble d'associations (je ne savais pas encore qu'on pouvait désigner ça comme des « réseaux d'affiliation »). […] Ces deux justifications sont inséparables de ma rencontre avec Nicolas Escach, dont j'avais lu la thèse (2014), et qui avait reconstruit empiriquement les « réseaux de villes » de l'espace baltique. Cela a été déterminant tant sur le plan méthodologique (sources, construction de bases de données, analyse de réseau) que sur les plans thématiques et théoriques.

LB : Dans votre cursus universitaire, avez-vous eu des cours d'analyse de réseau ? Si oui, quel type d'analyse de réseau (réseaux spatiaux, réseaux sociaux, autres) ? Si non, comment vous êtes-vous formées à l'analyse de réseau ?

EA : Je n'ai pas eu de cours « pratiques » sur l'analyse de réseau durant mon cursus universitaire (L1-M2). Néanmoins, nous avions vu le concept de « réseau » *via* des cours magistraux et TD, par exemple l'analyse des flux ou les infrastructures de transport dont les nœuds pouvaient structurer un territoire donné. Je me suis formée à l'analyse de réseau au début de la thèse à l'aide de collègues du laboratoire, puis en lisant de la littérature (ouvrages, articles scientifiques, sites en ligne) et en pratiquant sur logiciel(s).

NB : Dans mon cursus universitaire, je n'ai jamais eu de cours d'analyse de réseau. Je me suis formée principalement en thèse, par le biais de MOOC (« Des réseaux pour comprendre le monde » organisé par l'Université Paris 1), de formations doctorales organisées par l'Université Lyon 1 et une université d'été sur l'analyse de réseau. Ces formations ont été complétées par la lecture de manuels, de cours en ligne, d'articles scientifiques et de blogs (GDR Analyse de réseau en SHS, groupe fmr…). Les formations étaient centrées sur le corpus théorique (les différentes mesures, les auteurs et courants scientifiques qui ont modelé l'analyse de réseau) mais aussi sur les logiciels permettant leurs applications.

PG : […] Pendant ma licence de géographie à Albi, je n'ai abordé presque aucune approche quantitative, je savais simplement qu'elles existaient en géographie. En arrivant en master à Paris 1, j'ai découvert l'analyse spatiale et c'est dans le cadre de cours de ce genre que j'ai entendu parler de l'AR pour la première fois. Je me souviens notamment de cours de Claude Grasland dans le M2 Géoprisme. S'il y avait d'autres types de réseaux (je me souviens des relations entre moines étudiées par Samuel F. Sampson dans les années 1960), la plupart des exemples renvoyaient à des réseaux « spatiaux », dans le cadre d'une thématique sur les transports. Plus tard, quand j'ai enseigné moi-même l'analyse spatiale en L3, cela a confirmé cette impression : à Paris 1, on abordait principalement des réseaux de transport, en présentant des indicateurs associés à des graphes planaires, ce que j'ai toujours trouvé un peu bizarre car beaucoup de réseaux étudiés dans les exercices (comme le métro) ne sont pas, à proprement parler, planaires (ou en tout cas, il faut les « rendre » planaires un peu

artificiellement). Bref, avant d'entrer en thèse, je connaissais un peu de vocabulaire basique pour décrire des graphes (nœuds et liens ou sommets et arêtes, densité, centralité, diamètre, etc.) et je savais que les méthodes d'analyse étaient mobilisées pour plein d'autres réseaux, notamment les « réseaux sociaux », sans les connaître précisément.

J'ai complété ma formation en même temps que j'apprenais à coder avec *R*, à l'aide de plusieurs ressources sur le web, mais aussi grâce à de nombreuses interactions avec des chercheuses et chercheurs de Géographie-cités ou du groupe fmr (synthèses sur HAL et blog). […]

LB : De nombreuses disciplines utilisent l'analyse de réseau et la bibliographie sur le sujet est énorme, quelles ont été vos portes d'entrée bibliographiques ?

EA : La bibliographie sur l'analyse de réseau est en effet énorme ! Pour en avoir les clefs « initiales », le vocabulaire, les mesures les plus courantes, j'ai consulté des ouvrages et articles de plusieurs auteurs tels que Pierrre Mercklé (2004), Emmanuel Lazega (1998) ainsi que des articles sur le blog du groupe fmr (articles de Laurent Beauguitte et de Matthieu Devrelle notamment). En parallèle, j'ai également consulté des articles qui mobilisent l'analyse de réseau dans le champ de la santé, ces derniers sont moins nombreux. Généralement, les analyses sont fondées sur de « grandes bases de données » (études à partir des données de l'assurance maladie par exemple) ; à l'inverse des miennes ! Cet ensemble de lectures ont été des portes pour saisir des socles pratiques et des retours d'expérience sur l'emploi de cette méthodologie d'analyse.

NB : Mes portes d'entrées bibliographiques ont principalement été géographiques et sociologiques. L'analyse des réseaux sociaux (la *Social network analysis* nord-américaine notamment) a été une porte d'entrée importante car de nombreuses mesures ont été construites et utilisées dans ces travaux, et j'ai assisté au colloque de la *Sunbelt*, qui découle de cette école de pensée. Par ailleurs, les modèles proposés par Barabási & Albert (*scale-free network*) et Watts & Strogatz (*small-world network*) ont également été très utiles pour ma thèse. Mais ce sont principalement les travaux de géographie qui m'ont marquée et qui m'ont permis de transposer l'analyse de réseau sur mon sujet de recherche, notamment car de nombreuses ressources et articles pédagogiques sont mises à disposition par les géographes français. Les travaux menés par les géographes du laboratoire Géographie-Cités ont été fortement mobilisés et ont imprégné ma recherche ; je pense notamment aux travaux de Natalia Zdanowska, de Nadine Cattan, de Céline Rozenblat, de Marion Maisonobe ou aux vôtres. À l'échelle internationale, je me suis beaucoup appuyée sur les travaux de Peter Taylor et du GAWC[1].

[1] Groupe de travail sur les villes globales créé par le géographe anglais Peter Taylor à la fin des années 1990 et toujours actif aujourd'hui (https://www.lboro.ac.uk/microsites/geography/gawc/index.html).

PG : C'est un peu trop chaotique pour répondre précisément, ce qui est sûr c'est que je me suis souvent perdu dans la bibliographie sur l'AR... De nombreuses interactions plus ou moins formelles avec les collègues m'ont permis de découvrir différents papiers ou champ de recherche mobilisant l'AR : par exemple les travaux de Marc Barthelemy et ses collègues autour des réseaux spatiaux que Thomas Louail m'a indiqués à plusieurs reprises. Ensuite, il y a aussi des effets de « centralité » liés au nombre de citations de certains ouvrages et au fait qu'ils sont donc présents dans la plupart des bibliographies – je pense par exemple au manuel de Wasserman et Faust (1994) […] dont j'ai également entendu parler en master puis grâce à des collègues, en début de thèse. Il y a aussi des effets de langue avec la découverte de quelques manuels de synthèse en français : je pense notamment à celui de Degenne et Forsé (2004) ou au « Que sais-je ? » d'Emmanuel Lazega (1998). Cela m'a introduit à plein de problématiques qui m'intéressaient (tension théorique entre acteur et structure), à un vocabulaire et des méthodes (quels types d'enquête faire ? quelles *variables relationnelles* étudier ?) au sein de ce qu'on peut appeler « l'analyse structurale des réseaux sociaux », elle-même liée à des approches en sociologie des organisations notamment. Enfin, la documentation des packages *R* mentionne souvent, pour chaque fonction, les papiers théoriques et méthodologiques qui fondent les indicateurs : c'est aussi par ce biais que j'ai découvert le rôle important des physiciens (je pense notamment à Mark Newman) et des informaticiens dans ce qu'ils nomment souvent la *Network Science*. Tout ça m'a progressivement déplacé hors de la géographie, en tout cas en ce qui concerne l'AR […].

LB : Quels logiciels avez-vous utilisés ? Comment les avez-vous choisis ? Comment vous êtes-vous formé.e.s à ces logiciels ?

EA : Au début, j'ai utilisé le logiciel *Cytoscape* pour l'analyse de réseau. Cela m'a permis de voir comment structurer les données et d'avoir une interface plus facile à employer pour effectuer des mesures et des analyses. Puis, très vite, j'ai utilisé le logiciel *R* et *RStudio*, notamment les *packages igraph* et *visNetwork* (visualisation interactive) pour l'analyse de réseaux et les *packages FactoMineR* et *factoextra* pour les méthodes statistiques multivariées (ACP, CAH). L'avantage de ce logiciel est multiple. Il permet de traiter un nombre considérable de données rapidement (une fois le code écrit), de pouvoir modifier les visualisations, et les ressources en ligne sont nombreuses. Lorsque des points de blocages sont rencontrés, c'est un avantage. Pour l'utilisation de ces deux logiciels, j'ai été conseillée par les collègues du laboratoire de recherche, ensuite il a fallu apprendre en lisant des ressources sur ces logiciels et les pratiquer.

NB : Le premier logiciel utilisé est *Gephi*, et le choix a été guidé par sa prise en main très facile. Je l'ai utilisé jusqu'à la fin de la thèse car je trouve les visualisations produites très bonnes, malgré de nombreux bugs et une capacité de calcul limitée. Je me suis formée par le biais de MOOC et de tutoriels en ligne. Je me suis ensuite formée à l'analyse de réseau sur *R*, et en particulier au *package igraph*, qui m'a permis de

mobiliser d'autres indicateurs et algorithmes, notamment pour la détection de communautés. Je me suis formée grâce à une école d'été proposée à Nice en 2019 puis à des recherches autonomes en ligne. Enfin, je me suis également formée à la visualisation de réseaux sur le logiciel *QGis* pour la visualisation spatiale des réseaux de coopérations entre villes, grâce à des tutoriels sur le site du groupe fmr.

PG : J'ai eu une première approche « clique bouton » avec le logiciel *Gephi*. Cela m'a beaucoup aidé, car le caractère interactif de l'interface est très utile pour l'exploration, pour « toucher et voir » les réseaux qu'on a construit. Au fur et à mesure, j'ai intégré l'AR à ma pratique de *R* qui est devenu mon outil principal pour toutes mes analyses quantitatives (préparation de mes bases de données, statistiques, analyse spatiale, analyse de réseau, analyse textuelle et cartographie). En termes de ressources, ma pratique de *R* a été facilitée par l'ouverture d'un compte sur la plateforme *Huma-Num*, ce qui m'a donné accès à un serveur pour faire mes calculs.

On pourrait avoir l'impression que je décris ici une pratique autodidacte mais ce n'est pas du tout le cas. Si je n'ai pas vraiment suivi de formations à proprement parler, j'ai eu la chance de bénéficier d'une aide quasi-quotidienne pour me former à *R* de la part de mes collègues à Géographie-cités : Robin Cura (alors lui aussi en thèse), Hadrien Commenges, Thomas Louail pour ne citer que ceux qui ont eu à supporter des montagnes de questions de ma part. Une fois maitrisés les rudiments du langage-programme *R*, il n'est pas très difficile d'intégrer de nouveaux types d'analyse car il existe de nombreux *packages* à disposition, avec des fonctions pour faire les calculs. Ici, je veux juste insister sur le rôle décisif du logiciel libre et de la culture de partage qui va avec : en plus des différents supports de documentation et des nombreux tutoriels en ligne, les forums d'échange comme *Stack Overflow* constituent des ressources extrêmement précieuses pour se former, poser des questions précises, voire y répondre lorsque l'on sait…

Enfin, ma participation au *groupe fmr* a permis plein d'échanges avec Marion Maisonobe, Serge Lhomme, Françoise Bahoken et toi. J'ai découvert plein de choses sur les questions théoriques et méthodologiques propres à l'analyse de réseau. Les ateliers sur *R* qu'on a organisés avec Marion [Maisonobe] m'ont fait beaucoup progresser : comme souvent, c'est en enseignant qu'on apprend le mieux.

En fin de thèse j'ai utilisé un peu la plateforme *Cortext Manager* pour l'analyse textuelle (notamment pour la détection de mots-clés et de concepts), qui propose aussi quelques options d'analyse et de visualisation de réseaux.

LB : Tout ou partie de vos données a été construite sous forme de graphe biparti : avez-vous envisagé d'autres types de formalisation (hypergraphe par exemple) ? Avez-vous envisagé d'autres types de traitement que l'analyse de réseau (analyse des correspondances multiples notamment) ?

EA : D'autres formes de formalisation n'ont pas été envisagées dans ce travail de thèse. Le fait de représenter les liens « d'un point A à un point B » était important pour ce

travail pour retracer d'une part des systèmes de relations collaboratifs entre les acteurs. Ces graphes, croisés à certaines cartographies, m'ont permis de mieux expliquer des effets de localisation d'acteurs. D'autre part, des « filières » de soins de santé - révélées par une typologie des caractéristiques des *ego-networks* des acteurs – ont pu être éclairées par des graphes produits sur la base des données obtenues auprès des personnes vivant dans les quartiers de bâti spontané.
Enfin, croiser les graphes fondés sur les données obtenues auprès des acteurs aux graphes fondés sur l'analyse des données obtenues auprès des populations migrantes des quartiers enquêtés m'a permis de mieux expliquer certaines positions d'acteurs (centralité d'intermédiarité, d'autorité et de *hub*) dans la prise en charge des soins de santé des populations. Ainsi, les différents graphes obtenus par l'étude des données, se sont éclairés les uns les autres. Étant donné que la méthodologie se fonde sur l'analyse de réseau tout au long de la thèse, d'autres traitements n'ont pas été envisagés.

NB : Je n'ai pas envisagé d'autres types de formalisation, principalement car je n'avais pas connaissance de ces dernières. Il s'agit cependant de pistes que j'aimerais explorer par la suite. J'ai par ailleurs fait le choix de faire des traitements statistiques à partir des résultats de l'analyse de réseau (degré, centralité d'intermédiarité pour les villes qui coopèrent) plutôt qu'à partir du nombre de projets ou d'associations auxquelles elles participent car le nombre de villes partenaires varie fortement selon le nombre de projets ou d'associations. Aussi, par rapport à d'autres types de traitement statistiques, l'analyse de réseau me permettait de rajouter de la finesse dans l'étude des phénomènes coopératifs.

PG : […] J'ai essayé de travailler au maximum sur les graphes bipartis, en réalisant relativement peu de projections (transformation du biparti en graphes villes-villes ou associations-associations). J'ai effectivement découvert au cours de ma thèse, grâce à toi d'ailleurs, la possibilité de ne pas s'en tenir aux seules relations dyadiques pour modéliser des relations d'ordre supérieur, notamment par le biais d'hypergraphes. Là, pour être honnête, cela m'a semblé être une porte d'entrée un peu plus délicate et trop complexe pour moi. Pour les associations, dont certaines comprenaient plusieurs milliers de villes membres, cela ne me semblait théoriquement pas très pertinent. En revanche, pour la logique de projet, cela semble être une modélisation très convaincante. Lorsque l'on dit que 4 villes « coopèrent » ensemble pendant trois ou quatre ans, et que l'on n'a pas accès aux relations effectives entre les individus qui font vivre le projet (visites d'étude, e-mails ou téléphone, présence aux réunions, etc.), cela fait complètement sens de concevoir un lien unique entre les 4 villes symbolisant la coopération et la production d'un rapport de projet en commun, plutôt que de supposer l'existence de liens réciproques entre chaque couple de villes […] sans les décrire plus avant. Donc non, je n'ai pas vraiment fait ça dans ma thèse, par manque de temps car j'ai découvert trop tard ces approches […]. Je range donc les

hypergraphes dans la case « perspectives stimulantes » pour la suite de mes recherches.

Pour les autres types d'analyses que celles relevant strictement de l'AR : oui je les ai envisagés, voire même testés. J'ai fait de nombreuses analyses statistiques pour comprendre les niveaux de participation des villes aux dynamiques de coopération et des analyses multivariées pour synthétiser des logiques de localisation ou d'affiliation selon des thématiques. Toutefois, dans la thèse, j'ai privilégié le corpus de l'AR pour analyser les données relationnelles car je voulais mettre en avant la structure des systèmes de relations ainsi que les différenciations plus locales (positions des villes et des organisations dans le graphe, centralités). Mon approche théorique sur la question de la circulation des politiques urbaines m'a aussi encouragé à formaliser ces réseaux sous la forme de canaux potentiels pour l'échange d'informations : la modélisation de réseaux en graphes me semble particulièrement adéquate pour faire cela.

LB : Le nombre de mesures et de méthodes disponibles en analyse de réseau est important : comment avez-vous choisi parmi les nombreuses options disponibles ? Y a-t-il des méthodes que vous n'avez pas réussi à utiliser ? Des mesures que vous n'êtes pas parvenu.e.s à interpréter ? Des traitements que vous envisagiez et que les logiciels utilisés ne permettaient pas de mettre en œuvre ?

EA : Je me souviens avoir fait un tableau recensant des mesures d'analyse de réseau en notifiant les avantages et les limites de leur(s) apport(s). En fonction des questionnements posés dans cette recherche, certaines mesures se sont révélées être adéquates plus que d'autres, c'est pourquoi elles ont été choisies. J'ai notamment mobilisé la centralité de degré pour repérer les acteurs les plus cités et la transitivité locale (*clustering coefficient*) pour repérer les acteurs présentant un niveau élevé de coopération pour la prise en charge des besoins des populations vulnérables. J'ai également calculé l'intermédiarité pour repérer les acteurs qui sont des points de passage incontournables et les mesures de *hub* et d'autorité. De nombreux traitements sont disponibles dans *R*, je ne me souviens pas avoir été limitée par ce dernier.

NB : J'ai choisi les mesures en fonction de celles qui revenaient le plus dans la littérature et dans les formations, parce qu'elles me semblaient en même temps les plus accessibles et compréhensibles à lire et à transmettre, mais aussi parce que leur utilisation était mieux expliquée sur les différents logiciels utilisés.

Je n'ai pas réussi à utiliser un indicateur, celui de *l'assortativité*. Je souhaitais savoir si des villes coopéraient avec des nœuds ayant un degré similaire. J'avais pu voir, au travers de tris croisés, que cela me semblait être le cas. Cependant, je n'ai jamais réussi à avoir un résultat probant, je ne sais pas si cela est dû à un mauvais calcul ou bien si je n'arrivais pas à interpréter le résultat proposé. Par ailleurs, j'étais intéressée par l'étude des *k-means* mais j'ai eu du mal à savoir en quoi elles différaient de la détection de communautés et je n'ai pas trouvé de ressources en ligne claire sur ce point.

Enfin, certains traitements étaient complexes avec *R* et *Gephi* avec un graphe aussi gros que celui sur les associations de villes (900 000 liens), et j'ai eu du mal à visualiser et réaliser certains traitements sur ce graphe.

PG : Sans même évoquer les différences de vocabulaire et de lexique, il y a effectivement une extraordinaire diversité de méthodes et d'indicateurs. J'ai essayé un certain nombre de choses. Comme je me formais en même temps que je concevais mes analyses, j'ai fait énormément d'explorations et je me suis vite retrouvé un peu perdu. À ce moment-là, j'ai eu besoin de formaliser une chane de traitements à appliquer sur mes deux grands types de réseaux, afin de ne pas partir dans tous les sens et de faciliter la comparaison. Avec les ressources en ligne et la bibliographie, deux choses me sont apparues. Premièrement, dans les différentes ressources liées à *R*, j'ai cru apercevoir des sortes de routines qui revenaient souvent : création du graphe (et projection si graphe biparti), calcul de quelques indicateurs globaux (taille, ordre, diamètre, composantes connexes, densité), calcul d'indicateurs locaux (principalement des indicateurs de centralité), détection de communauté (souvent avec l'algorithme de Louvain). Dans beaucoup d'articles en revanche, on trouvait des indicateurs ou méthodes spécifiquement développés ou adaptés au cas d'étude. J'ai essayé de me situer un peu entre ces deux tendances : j'ai donc fait une chaîne de traitements qui reprend les grandes étapes des routines observées sur internet, en allant du plus simple au plus complexe, mais avec une attention particulière à la signification de chaque indicateur et mesure dans mon cas d'étude. Avec ce travail, j'ai écarté tout ce que je ne pouvais pas interpréter et j'ai aussi fait des choses un peu plus « originales », comme aborder la centralité des villes par les caractéristiques de leurs *ego-networks* sous forme bipartie (plutôt que de m'en tenir aux seuls indicateurs de degré ou d'intermédiarité).

De fait, en enquêtant un peu plus sur les méthodes concernant les graphes bipartis, j'ai dû renoncer à appliquer certains algorithmes qui me semblaient théoriquement intéressants mais que j'étais incapable de coder sous *R*.

Une dernière chose sur les renoncements (j'ai déjà évoqué celui des hypergraphes) : mon approche de l'AR a principalement été descriptive. J'ai découvert plusieurs approches de modélisations / simulations, par exemple les *ERGM – Exponential Random Graph Modelling* – souvent mobilisés en sociologie. Mais je n'avais pas tous les éléments et compétences pour mobiliser ces approches sereinement. *A posteriori* cela m'apparait très utile, notamment pour répondre à des questions sur les processus d'association en distinguant les effets propres à la structure du réseau de ceux qui dépendent de variables inhérentes aux acteurs. Les modélisations qui comparent le réseau empirique à des graphes générés aléatoirement m'auraient également permis de mieux prendre en compte la dimension temporelle et évolutive.

Figure 1 : De la récolte des données au choix des indicateurs

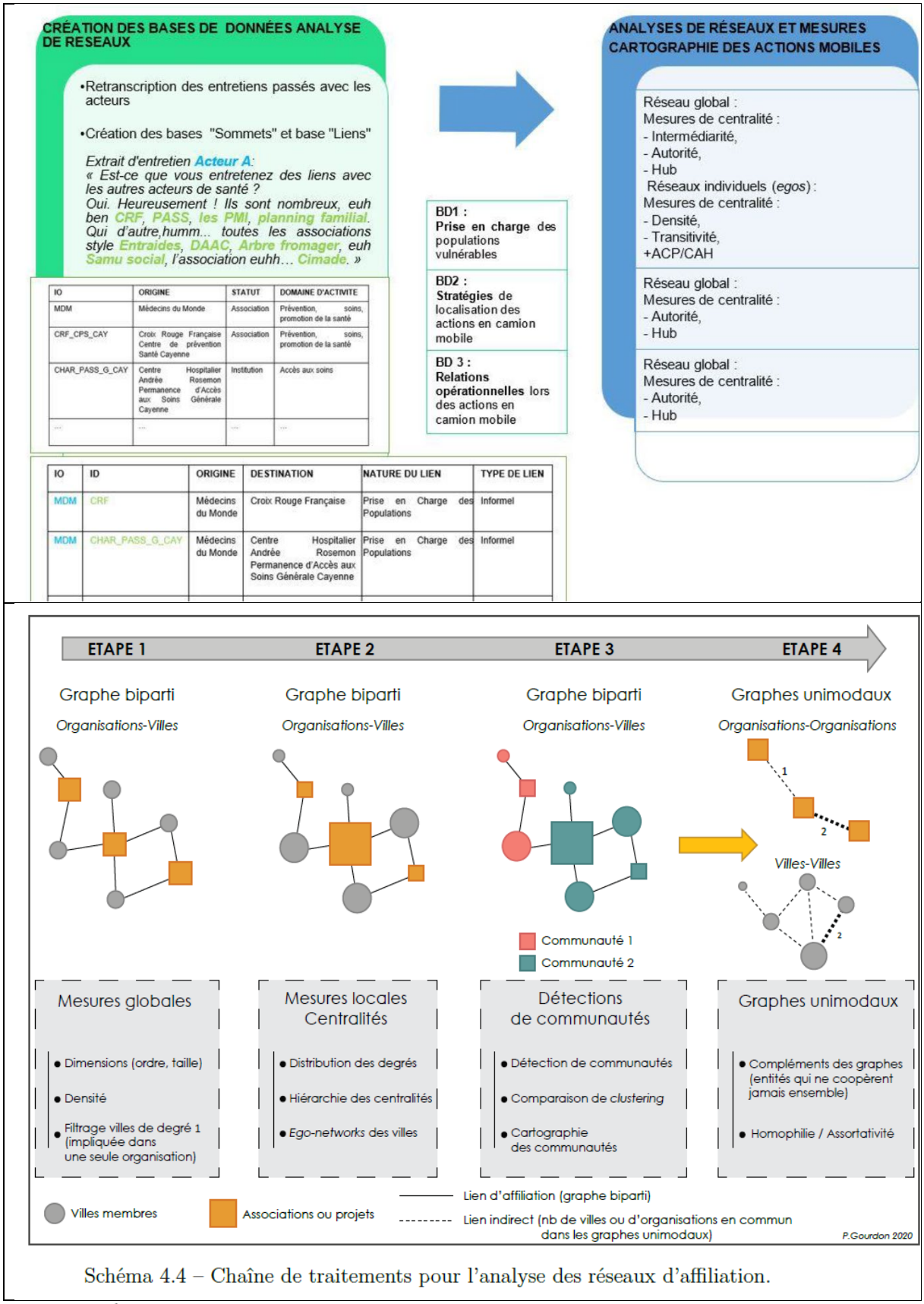

Schéma 4.4 – Chaîne de traitements pour l'analyse des réseaux d'affiliation.

En haut : Élise Autrive, 2021, p. 278. En bas, Paul Gourdon, 2021, p. 262.

LB : Concernant la visualisation des données relationnelles, vous avez privilégié la forme dite lien-nœud (sommets et liens) : cette forme vous paraît-elle toujours satisfaisante ? Avez-vous envisagé d'autres formes de visualisation ?

EA : Cette visualisation « liens-nœuds », qui comporte des avantages dans ce travail de thèse, comporte aussi des limites ! […] Typiquement le fait de représenter au centre les acteurs les plus connectés et en périphérie les acteurs les moins connectés peut masquer certains nœuds périphériques qui ont eux aussi leur importance. Par exemple certains acteurs se situaient en périphérie, pour autant ils étaient importants si l'on se place du point de vue de la prise en charge de besoins spécifiques de certains groupes de populations. Ainsi il est important d'avoir accès à des dimensions plus qualitatives et de contexte sur les données au risque de masquer certaines parties interprétatives de la structuration du graphe.

Par ailleurs, la forme « liens-nœuds » peut être malaisée en cas de graphes avec une multitude de nœuds et de liens qui s'entrecroisent dans la mesure où le système de relation devient difficile à interpréter avec une seule lecture visuelle. Dans le cadre de travail le plus grand graphe était composé de 91 nœuds ; je ne me suis donc pas vraiment heurtée à ce souci et n'ai pas eu à trouver d'autres formes de visualisation. Enfin, il est parfois difficile de faire en sorte que des liens ne se croisent pas entre les nœuds, ceci pouvant entraver la lecture.

NB : J'ai été réellement bloquée dans l'analyse de mon deuxième réseau, pour les raisons expliquées au-dessus, aussi la forme lien-nœud me semble limitée à partir d'un certain nombre de liens. Je n'avais cependant pas réellement entendu parler d'autres formes de visualisation avant le rendu de ma thèse. Depuis, j'ai entendu parler de formes permettant de regrouper les nœuds et de diminuer ainsi le nombre de liens, mais je n'ai pas encore étudié cette méthode, qui pourrait être prometteuse.

Figure 2 : Modèles graphiques

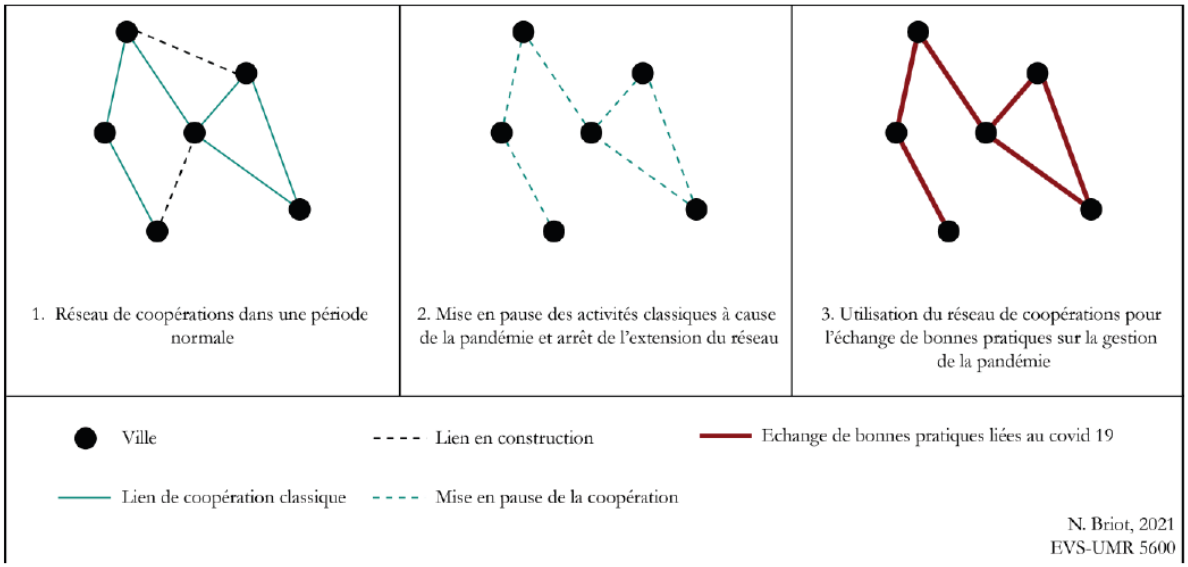

Figure 7.8 : Le réseau de coopérations entre villes pendant la pandémie

Ninon Briot, 2021, p. 396

PG : Au risque de répondre un peu à côté de la plaque, je pense que pour la question de la visualisation, il est nécessaire de replacer chaque image produite dans un tout, une démarche d'ensemble. Quand j'y pense, connaitre le diamètre d'un graphe n'est pas une visualisation mais, pour moi, ça produit une figuration importante du réseau que j'étudie. De fait, dans la thèse, j'ai produit pas mal de visualisations différentes, dont certaines s'écartent de la représentation de graphes sous la forme liens-nœuds. L'idée c'est de combiner plusieurs possibilités pour présenter des résultats de méthodes d'AR (distribution des indicateurs de centralités, cartographies de communautés, etc.). Étant donné les dimensions des réseaux étudiés, je n'en ai représenté aucun directement car cela aurait produit des « toiles d'araignée » illisibles. […] J'ai l'impression que, quel que soit le mode de représentation (lien-nœud, *heatmap*, diagramme en arcs, etc.), dès lors qu'il y a plusieurs milliers de sommets et de liens, il n'y a pas de « bonnes » représentations (statiques). Cela donnera simplement cette impression de complexité et de polarité mais le graphique produit ne pourra pas être lu comme une source d'information (détaillant l'existence de chaque relation). Cette illisibilité relative me semble une bonne occasion de produire de l'analyse, de la synthèse, et d'explorer d'autres modes de représentation qui ne soient pas directement relationnels. Dans ces cas-là, le calcul d'indicateurs et les filtrages successifs sont des étapes cruciales pour décomposer l'ensemble en parties plus facilement représentables. D'autant que ramener des structures complexes et foisonnantes à quelques motifs/schémas relationnels très simples est particulièrement fécond pour l'interprétation et la généralisation théorique. Les représentations des données issues de ce travail, bien qu'incomplètes, deviennent alors beaucoup plus intéressantes du point de vue du message qui est transmis par la visualisation.

Je me suis donc appuyé sur beaucoup d'autres représentations pour figurer les résultats des analyses. À titre d'exemple, j'ai utilisé des représentations *Rank Clock* pour mettre en avant la variabilité des hiérarchies de villes selon les indicateurs ou filtrages des réseaux considérés. Autre exemple : pour représenter les communautés de villes telles que détectées dans le réseau, j'ai exploré d'autres possibilités que la carte « *smarties* » (un semis de points où chaque ville a une couleur en fonction de sa communauté d'appartenance). Dans ce dernier cas, je me suis dit : « le but de la carte n'est pas de donner une information précise sur la communauté d'appartenance de plusieurs milliers de villes mais de s'interroger sur les formes spatiales (emprise, extensions, polarités) du semis de points de chaque communauté », j'ai donc produit une carte « météorologique » dans laquelle les semis des communautés sont transformés en isolignes concentriques (grâce au calcul des densités de Kernel). Toutefois, il est vrai que pour la dimension relationnelle, j'ai principalement représenté quelques graphes bipartis et certaines projections (villes-villes) mais, à chaque fois, avec des filtrages drastiques sur le nombre de liens et de nœuds affichés. On peut résumer ainsi mon approche : pour la description des grandes structures, recourir au calcul d'indicateurs globaux et locaux ; pour figurer la dimension relationnelle, procéder au filtrage du graphe afin de répondre à une question particulière, assortie d'une visualisation liens-nœuds (avec légende et résumé statistique).

LB : Vous avez tout.e.s les trois réalisé des entretiens : avez-vous eu l'opportunité de montrer vos images de réseaux aux personnes enquêtées ?

EA : Les images de réseaux produites ont pu être consultées par un petit ensemble de personnes avec qui je suis restée en relation et que j'ai invitée à ma soutenance de thèse (notamment les acteurs « de terrain ») mais elles n'ont pas commenté ces figures. Sinon, toutes les personnes interrogées n'ont pas vu les réseaux produits, je pense notamment aux personnes en situation de migration enquêtées dans les quartiers.

NB : Je n'ai pas eu l'occasion de montrer les images de mes réseaux aux enquêtés alors que mes entretiens se basaient sur les résultats des traitements de l'analyse de graphe (degrés, intermédiarité, villes les plus connectées, *clusters* détectés). J'envisage aujourd'hui de produire un document synthétique à destination des acteurs interrogés, et je me pose de nombreuses questions sur la manière de rendre ces analyses intelligibles, mais surtout utiles pour la mise en place de politiques de coopération. Une idée serait de produire une sorte d'état des lieux de l'ego-réseau de chaque ville interrogée, en le croisant avec ce qu'elles ont pu me dire en entretien, mais cela prendrait beaucoup de temps.

PG : Non pas vraiment… J'ai seulement mentionné des résultats ou évoqué, en fin d'entretien, des questionnements sur des liens récurrents que j'avais repéré grâce à l'AR. Cependant, au moment où j'ai mené la plupart des entretiens, je n'avais pas encore fini mon travail sur les bases de données et leurs explorations.

LB : Plus généralement, quels ont été pour votre travail les principaux avantages et les principales limites de l'analyse de réseau ?

EA : La mobilisation de l'analyse de réseau pour étudier les données obtenues par les entretiens et les questionnaires constitue une plus-value pour l'étude de ce sujet de thèse dans la mesure où elle a permis d'une part de faciliter la représentation des différentes relations : entre acteurs, entre lieux de recours et individus (selon différents types de soins de santé). D'autre part, par la mise en place de mesures (centralités d'intermédiarité, d'autorité, de *hub*, transitivité), elle a permis d'approfondir les interactions des acteurs en révélant des fonctions et rôles spécifiques d'acteurs dans la structuration des réseaux.

Concernant l'Agence Régionale de Santé (ARS) et ses différents pôles, ces analyses ont permis d'identifier que leurs visions respectives sur la structuration des acteurs y est différente. Concernant les acteurs des soins de santé, leurs « relations de collaboration en partenariat » étudiées d'un point de vue global ont permis de dévoiler une *structuration informelle* (sans conventions) des acteurs dans la prise en charge des populations. Se recentrer ensuite sur les réseaux individuels des acteurs (ego-réseaux) a finalement permis de faire émerger une typologie des acteurs laissant apparaître des formes de coordination entre acteurs de différents groupes, mais aussi des mailles d'intervention des acteurs communes aux acteurs des groupes établis. Concernant les populations, ces analyses de réseau ont permis d'identifier des lieux de recours locaux pratiqués des populations pour différents types de soins de santé (vaccination, dépistage, accompagnement social) et d'aides à caractère social (aide alimentaire, aide vestimentaire). Par exemple, pour se faire vacciner, le Centre de Prévention de Santé de la Croix Rouge Française en Guyane a été identifié comme le principal lieu de recours des personnes interrogées. Cet acteur avait par ailleurs été identifié comme « point de passage important » (centralité d'intermédiarité) dans la prise en charge des populations vulnérables à partir des analyses portant sur les entretiens effectués avec les acteurs.

L'analyse de réseau a permis de mettre en lumière l'importance de prendre en compte les liens (bien qu'informels) dans la structuration des espaces d'organisation des acteurs et des populations dans le champ de la santé ; alors que ces liens (informels) n'apparaissent pas toujours dans les critères d'évaluation (normés par des conventions de partenariat par exemple) des acteurs.

L'une des limites est que la structuration d'un réseau ne peut à elle seule tout expliquer ! D'où l'importance d'avoir également des données qualitatives et de contexte.

Croiser ces résultats aux analyses sémantiques des entretiens et aux cartographies réalisées a permis de mieux interpréter les mesures effectuées et renforcer la compréhension de la structuration des réseaux, soit parce que ces matériaux constituaient « en amont » une hypothèse pour formaliser l'analyse du réseau ; soit parce qu'ils permettaient d'approfondir « en aval » l'analyse du réseau.

Ces matériaux recueillis sur le terrain ont ainsi été essentiels pour mener cette étude et ont permis d'éclairer et de conforter l'existence de structurations non visibles, c'est-à-dire établies sur des liens informels entre les acteurs. Menée sur ces différents acteurs et les populations, ces analyses de réseau ont en somme facilité la compréhension de la structuration globale des acteurs investis dans la prise en charge des populations vulnérables telle qu'elle est vue (ARS), exercée (acteurs des soins de santé) et pratiquée (populations).

NB : Le principal avantage de l'analyse de réseau est qu'elle m'a donné les outils nécessaires à la visualisation et la compréhension d'un réseau de relations peu étudié par la littérature scientifique, en l'abordant par le prisme de sa structure. Cette visualisation m'a permis de faire le pont entre une littérature très tournée vers la science politique, et concentrée autour de cas précis de villes internationales ou de projets de coopérations, à la géographie, et notamment aux impacts spatiaux de ces relations.

Ce corpus méthodologique contraint, de plus, à une réelle réflexion sur les données, sur la pertinence de certaines mesures lorsque l'on n'a pas un graphe complet, et m'a donc forcé à produire un recensement assez proche de la réalité de l'état des coopérations entre villes.

Les indicateurs de l'analyse de réseau m'ont enfin été très utiles pour comprendre la réalité des relations entre villes. En effet, les mesures et le corpus notionnel permettent de repenser ce qu'est la centralité d'un nœud. La notion d'intermédiarité, que je ne connaissais pas avant de commencer ma thèse, s'est révélée être une notion centrale pour la compréhension des phénomènes coopératifs. Pour être internationale, une ville doit certes coopérer beaucoup, mais elle devient aussi centrale quand elle permet de relier des ensembles qui ne le sont pas spécifiquement. À ce titre, on peut analyser la place de certaines villes complètement différemment, comme le cas des villes turques ou du Moyen-Orient, qui se placent comme relais entre villes occidentales et africaines/asiatiques. Aussi, derrière les méthodes et mesures, c'est aussi une nouvelle vision et conceptualisation des relations entre espaces que j'ai appris à maîtriser.

Cependant, l'analyse de réseau comporte certaines limites. Ces méthodes sont rarement le centre des formations en sciences sociales, aussi leurs appréhensions sont parfois complexes en même temps pour celle ou celui qui les produit, mais également pour les personnes qui les lisent. Il est parfois dur d'arriver à simplifier ses résultats pour faire une communication, un article ou un cours.

Par ailleurs, si les notions et méthodes de bases sont très accessibles en ligne et l'on se forme assez rapidement aux rudiments de l'analyse de réseau, il est beaucoup plus compliqué de se former en profondeur à ce corpus méthodologique, tant sur le plan pratique (maîtrise de méthodes un peu originales, indicateurs peu connus, logiciels ou *packages* dans toutes leurs possibilités) que conceptuel. En effet, le plus dur pour moi a été d'avoir un œil critique sur les mesures que je pouvais produire, sur le choix des indicateurs pour la démonstration… Les méthodes de détection de

communautés, par exemple, sont très variées, et il n'est pas aisé de savoir laquelle choisir. Il en va de même avec les mesures de centralité (j'ai choisi d'utiliser le binôme assez classique degré-intermédiarité, mais je n'ai pas su quoi faire des autres formes de centralités comme la *closeness*, *eigenvector*…). Aussi, si l'analyse de réseau permet de produire des recherches originales en sciences sociales, c'est également cette originalité qui peut parfois être une limite, dans sa production comme sa diffusion.

PG : J'ai un peu de mal à voir des avantages ou limites dans l'absolu. Je dirais que ça dépend de ce qu'on cherche à faire. Pour moi, les limites sont toujours extrêmement importantes à concevoir et sont au centre de chaque commentaire, interprétation, analyse (on m'a souvent dit d'ailleurs que je passais trop de temps à décrire ces limites et à dire ce que je n'avais pas fait, c'est vrai que ce n'est pas forcément très « vendeur »). Je tiens beaucoup à l'idée que chaque calcul ou visualisation participe à la construction d'un problème et appelle de nouvelles questions : ça montre quelque chose mais ça montre également « ce qui manque ». Pour moi, l'analyse de réseau a donc été très utile car prenant place dans une démarche exploratoire des données plus large et structurée en trois temps : d'abord l'analyse des localisations (cartographie, analyse spatiale) ; ensuite l'analyse des données relationnelles (analyse de réseau) ; et enfin analyse des contenus politiques (discours, modèles urbains) qui circulent dans les canaux (décrits dans les deux étapes précédentes). L'analyse de réseau est donc principalement intervenue dans les phases 2 et 3 de ma démarche exploratoire. Pour la phase 2, c'est assez évident. Pour la phase 3 (sur les contenus politiques), l'AR est mobilisée ponctuellement avec d'autres outils de l'analyse textuelle, afin d'explorer des liens entre catégories et labels de l'action publique. Il s'agissait alors de figurer un lexique, un espace sémantique, afin d'objectiver la manière dont des pratiques locales (politiques urbaines) étaient « traduites » en « bonnes pratiques », dessinant ainsi une sorte de répertoire des politiques urbaines au niveau européen.

Pour finir, je voudrais formuler une remarque (plus qu'une limite) liée à ma pratique de l'AR dans le cadre de ma thèse. Dans la plupart des articles que l'on peut lire, la dimension recherche et exploration sur les données est en quelque sorte gommée. On présente des résultats. Lorsqu'on n'a pas encore conduit soi-même des analyses, on est alors aisément fasciné de voir comment les méthodes de l'AR semblent puissantes, montrent plein de choses. Lorsque l'on est face à ses propres réseaux, c'est un peu une autre histoire… On se rend compte qu'il faut faire un nombre considérable d'analyses et de visualisations pour commencer à se familiariser avec l'objet, pour lui poser des questions intéressantes. Les premières analyses canoniques (indicateurs globaux, détections de communauté) peuvent se révéler décevantes par rapport à ce qu'on a lu, bien présenté, dans les articles scientifiques : ou bien c'est parfaitement trivial, ou bien ça semble ne rien montrer du tout… […] Il faut du temps, assumer le caractère construit et partiel de tous les choix de modélisation, de tous les filtrages et là, on peut commencer à répondre à des questions et, surtout, à en poser d'autres.

Références des trois thèses

Élise Autrive, 2021, *Territorialisation sanitaire et accès aux soins de santé primaires pour les populations vulnérables de la communauté d'agglomération de Cayenne (Guyane Française): une approche par l'analyse de réseau*. Thèse de doctorat. Université de Normandie. En ligne.

Ninon Briot, 2021, *Villes en réseaux. Les coopérations internationales des villes françaises : spatialisation, internationalisation, européanisation*. Thèse de doctorat. Université de Lyon. En ligne.

Paul Gourdon, 2021, *La coopération entre villes européennes: convergences dans l'action publique urbaine par la circulation transnationale de modèles*. Thèse de doctorat. Université Paris 1. En ligne.

Références citées dans l'entretien

Alain Degenne et Michel Forsé, 2004, *Les réseaux sociaux*, Armand Colin.
Nicolas Escach, 2014, *Réseaux de villes et recompositions interterritoriales dans l'espace baltique*, Thèse de doctorat. Géographie, ENS Lyon. En ligne.
Emmanuel Lazega, 1998. *Réseaux sociaux et structures relationnelles*, Presses universitaires de France.
Patrick Le Galès, 2011, *Le retour des villes européennes,* Presses de Sciences-Po.
Pierre Mercklé, 2004, *La sociologie des réseaux sociaux*, La Découverte.
Stanley Wasserman et Katherine Faust, 1994, *Social Network Analysis. Methods and Applications,* Cambridge University Press.