



HAL
open science

Sources et données pour l'analyse des réseaux sociaux

Marie-Pierre Renée Bes, Guillaume Favre, Claire Lemerrier

► **To cite this version:**

Marie-Pierre Renée Bes, Guillaume Favre, Claire Lemerrier. Sources et données pour l'analyse des réseaux sociaux. *Bulletin de Méthodologie Sociologique / Bulletin of Sociological Methodology*, 2021, 152 (1), pp.10-51. 10.1177/07591063211040228 . halshs-03413207

HAL Id: halshs-03413207

<https://shs.hal.science/halshs-03413207>

Submitted on 3 Nov 2021

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.

Sources et données pour l'analyse des réseaux sociaux

Marie-Pierre Bès

LISST, Université Toulouse Jean Jaurès, Toulouse, France

Guillaume Favre

LISST, Université Toulouse Jean Jaurès, Toulouse, France

Claire Lemerrier

Centre de sociologie des organisations, CNRS, Paris, France

Corresponding Author

Claire Lemerrier, claire.lemerrier@sciencespo.fr, CSO, 19 rue Amélie, 75007 Paris, France.

Abstract: Sources and data for social network analysis

Social network analysis (SNA) is becoming more and more widespread in several scientific disciplines as a method for processing social, economic, geographic, historical, digital, etc. data. Visualizations of graphs, communities and other ties are multiplying. However, there is a risk that SNA users may lose sight of the conditions under which their data are produced and thereby overinterpret or underinterpret the results. This article provides a guide to the different types of sources of network data, and the pitfalls and choices encountered in the process of producing network data. The first part provides an overview of these pitfalls and choices, and the second part focuses on the specifics of each source of data. Based on the authors' experience in training network researchers, the article proposes a review of controlled ways of producing network data from this array of sources.

Résumé

L'analyse des réseaux sociaux (ARS) ne cesse de se répandre dans plusieurs disciplines scientifiques comme une méthode de traitement de données sociales, économiques, géographiques, historiques, numériques, etc. Les visualisations de graphes, communautés et autres liens se multiplient et le risque existe que les utilisateurs et utilisatrices de l'ARS perdent de vue les conditions de production de leurs données et surinterprètent ou sous-interprètent de ce fait les résultats. Cet article propose un guide des différents types de sources qui se prêtent à l'ARS, en mettant l'accent sur les opérations de transformation de ces sources en données de réseau, les écueils qu'elles présentent et les grands arbitrages qu'elles impliquent. La première partie donne une présentation générale de ces éléments, que la seconde partie décline par type de source. Fondé sur une expérience de formateur et formatrices, l'article propose donc une recension des façons de produire des données de réseau en contrôlant au mieux leur rapport aux sources.

Keywords : Social network analysis, data, sources, source criticism, name generators, digital trace data, sociometric questionnaires

Mots clés : Analyse des réseaux sociaux, données, sources, critique des sources, générateurs de noms, traces numériques, questionnaires sociométriques

Bien au-delà de l'ordre alphabétique, Marie-Pierre Bès est la première auteure de cet article, né d'une intervention à trois en formation sous le soleil de Cargèse. Elle nous a poussés inlassablement à en tirer une version écrite et en a elle-même composé une grande partie. Marie-Pierre a été emportée durant la phase de finalisation de l'écriture. Cet article témoigne de son engagement comme enseignante et comme chercheuse en ARS, soucieuse du travail bien fait et marquante par son inventivité. Ce que nous retiendrons de Marie-Pierre est qu'elle était toujours capable de formuler des questionnements sociologiques originaux et des protocoles d'enquête et de collecte de données en partant d'interrogations en apparence naïves. Ses travaux et réflexions sur les chaînes relationnelles et les inégalités mériteraient d'être prolongés, nous espérons que cet article y contribuera.

Introduction

Le vocabulaire des « réseaux » et les visualisations en réseaux ne cessent de gagner en popularité depuis plus d'une vingtaine d'années. Dès les années 1990, la tradition déjà ancienne en anthropologie et sociologie d'« analyse des réseaux sociaux » (ARS – à son sujet, voir Freeman, 2004)¹ avait donné lieu à des publications de plus en plus nombreuses sur des thèmes variés, sans doute à la fois grâce à la plus grande accessibilité des outils informatiques et du fait d'un intérêt croissant pour les liens sociaux, le capital social ou encore les relations entre organisations. À l'origine, l'ARS s'inscrit dans des démarches théoriques claires et relativement bien délimitées, celles de l'anthropologie structuro-fonctionnaliste ou d'une sociologie structurale (Knox et alii, 2006) par exemple. Les concepts, méthodes et indicateurs utilisés sont ancrés dans des théories et permettent de répondre à des problématiques relativement bien définies : connectivité du monde social, sociabilité, rôles sociaux, etc. Dans les années 2000, l'ARS a pris encore plus d'ampleur mais a aussi perdu en unité : d'abord lorsque que des chercheur·ses venu·es de la physique ont commencé à analyser des réseaux de relations entre êtres humains, en employant un vocabulaire différent (des différences bien résumées par Borgatti et alii, 2009) ; ensuite du fait de la montée en puissance des « réseaux sociaux » numériques (Facebook, puis Twitter et bien d'autres) et des études à leur sujet (voir Mercklé, 2016, dont les rééditions successives témoignent d'ailleurs de ces évolutions). L'ARS est petit à petit devenue une méthode générique d'analyse de données, notamment de données nativement numériques. Elle est en outre toujours plus accessible : on peut rapidement obtenir en ligne des visualisations « en réseaux » de n'importe quelles données et les colorer pour y mettre en évidence des « communautés ».

Mais cela ne fournit pas clés en mains le moyen d'interpréter ces graphes : qu'en dire, au-delà du fait qu'ils sont compliqués, ou au contraire qu'on y retrouve de grandes structures bien connues ? Ce texte est né de notre expérience de formateur et formatrices en analyse de réseaux sociaux. Nous avons constaté au fil des années que l'essentiel n'était pas d'expliquer le fonctionnement des logiciels permettant de visualiser des graphes ou de calculer des indicateurs. Pour que les personnes qui se forment puissent, chacune dans leur discipline, avec leurs questions théoriques et leurs matériaux empiriques chaque fois particuliers, tirer une réelle valeur ajoutée de l'utilisation de ces logiciels, il faut, dans notre expérience, deux ingrédients

¹ Nous utiliserons dans le reste de l'article le sigle ARS pour décrire l'analyse de relations pertinentes pour les sciences humaines et sociales, qu'il s'agisse ou non de relations entre humain·es, que ces relations utilisent ou non un support numérique, quelles que soient la discipline de rattachement de l'analyste et la méthode employée.

supplémentaires : une réflexion sur le sens du vocabulaire propre à l'ARS dans le cas étudié et, plus important encore, sur la manière de construire des données adaptées à chaque recherche.

Le premier est fourni par les manuels existants en ARS : il s'agit de réfléchir à ce que peuvent signifier différentes structures, micro-structures ou positions repérables dans les réseaux, par exemple le fait que l'on trouve de l'« homophilie » dans un réseau, ou encore de la « transitivité », ou que l'« intermédialité » d'un individu est élevée. Il est toujours bon de transmettre ce vocabulaire spécifique, et son ancrage dans des théories, de manière interpersonnelle plutôt que générique, en recherchant avec les personnes en formation quel indicateur pourrait correspondre à leurs questions ; mais les manuels fournissent une base précieuse².

Ils sont en revanche beaucoup moins complets, en général, sur le second élément (cf. Tableau 2), ce qui nous a donné envie de lui consacrer cet article. Nous avons régulièrement rencontré en formation des personnes qui pensaient que l'« analyse de réseaux » était une méthode pour *détecter* des liens sociaux sur leurs terrains d'enquête ou dans leurs archives. Ces personnes étaient très déçues lorsqu'elles se rendaient compte que beaucoup de présentations standards de la méthode commençaient avec des données de réseau déjà recueillies, mises en forme et, généralement, dans les exemples utilisés pour présenter les indicateurs, très simples : une relation étudiée au sein d'un groupe assez petit d'individus, pour lesquels on sait avec certitude qui a ou non un lien avec qui. Nous-mêmes, par nos recherches personnelles, savions que l'essentiel du temps se passait en réalité en amont, à construire ce type de données (et, souvent, à les simplifier) pour en arriver là ; et que bien des décisions déterminantes pour les interprétations futures sont en réalité prises pendant ce temps de la construction des données. Qu'il s'agisse de *big data* ou de l'étude d'un tout petit groupe – ce n'est d'ailleurs pas contradictoire –, les opérations de construction des données sont aussi cruciales pour l'ARS que pour les sciences sociales en général, mais elles présentent des spécificités qui sont rarement documentées dans les manuels, moins en tout cas que celles de l'analyse. C'est pourquoi nous lui avons donné de plus en plus de place dans nos formations et nous avons eu envie de résumer ici ce que nous en avons appris.

La première partie de cet article présente quelques principes de base. Ils ne constituent pas une méthodologie standard, mais plutôt un ensemble de questions à ne pas oublier de se poser explicitement, pour éviter les choix par défaut. En substance, chaque principe propose une alternative entre une stratégie de construction des données qui prend moins de temps et une autre qui produit des données plus riches, donc plus interprétables. Ces arbitrages se retrouvent bien sûr dans toute recherche, mais nous nous proposons d'explicitement quelles peuvent être ces sources de richesse interprétative en matière d'ARS, afin d'aider à faire ces choix. La seconde partie de l'article propose un panorama des différentes sources possibles de données pour l'ARS. Elle vise avant tout à souligner que l'espace des possibles est grand, loin de se limiter à ce dont les manuels ont longtemps parlé en priorité (l'enquête par questionnaire) ou à ce qui est à la mode (l'extraction massive de données de plates-formes numériques). Entretiens, observations, expérimentations, sources écrites peuvent aussi être utilisées pour construire des données de réseaux. Pour chaque grand type de source, nous signalons, en partie sur la base de la littérature, mais plus encore de nos expériences directes de recherche et de questions

² Nous renvoyons au Tableau 2 ci-dessous pour une liste des manuels en français et en anglais les plus célèbres dans la tradition sociologique de l'ARS. Un des effets de la convergence d'intérêts pour les réseaux sociaux entre disciplines depuis les années 2000 a été une diversification des vocabulaires qui impose de constantes opérations de traduction : les physicien·nes ont ainsi des indicateurs en partie différents de ceux des sociologues et appellent souvent différemment les mêmes indicateurs – et ainsi de suite. Une bonne manière d'appréhender les convergences et divergences entre vocabulaires est de se référer aux pages de la version anglaise de Wikipedia consacrées d'un côté à « Social network », de l'autre à « Network science ».

entendues en formation, quelques difficultés classiques, dans l'optique d'éviter aux débutant-es de se confronter à des problèmes déjà bien connus – ou plutôt de leur permettre de le faire en connaissance de cause (cf. Tableau 1).

Cet article va ainsi parcourir successivement les lignes, puis les colonnes du Tableau 1 (ce que nous n'avons jamais le temps de faire complètement en formation : nous incitons les lecteurs et lectrices à le lire sélectivement, ou en plusieurs fois, selon leurs besoins). Les lignes résument les différentes manières d'enrichir les données, lors de leur construction, que nous abordons successivement dans la première partie de cet article. Il peut s'agir d'une part de recueillir de l'information sur les conditions de production des sources elles-mêmes (et de documenter la construction qu'on opère ensuite à partir de ces sources), ce qui peut être important pour plusieurs raisons, que nous développons dans les points 1., 2. et 3. de la première partie. Ce recueil d'informations est nécessaire quelle que soit la source, mais capital lorsqu'on travaille à partir de matériaux préexistants : documents écrits, traces numériques ou encore questionnaires que l'on n'a pas produits soi-même. D'autre part, une seconde manière d'enrichir les données est de s'astreindre à produire autre chose qu'une « simple » liste de liens attestés (ce qui n'est déjà, en pratique, pas si simple !), en travaillant activement à documenter le statut des liens manquants, à rechercher des types de liens non triviaux, à qualifier précisément ces liens, mais aussi les entités qu'ils relient – ce que nous développons dans les points 4., 5., 6. et 7. Là aussi, cet enrichissement peut être pratiqué quelle que soit la source de données ; mais il est crucial, cette fois, lorsque celles-ci sont construites au moment de l'enquête (qu'elle ait lieu par questionnaire, entretien, expérimentation ou observation) et qu'il est très coûteux, voire impossible, de revenir ultérieurement sur son terrain pour compléter une première collecte.

Il est bien sûr rarement envisageable de construire des données riches sur tous ces plans à la fois : pas seulement parce que le temps d'une recherche est limité, mais aussi parce que certaines sources de données, quels que soient les efforts de construction, rendent telle ou telle forme d'enrichissement difficile. C'est ce que signalent les indications dans les cases du Tableau 1, de manière évidemment très simplifiée. Nous développons ces indications dans la seconde partie de cet article, qui suit cette fois les colonnes du Tableau 1, d'un grand type de source de données à l'autre. Nous y traitons toutefois ensemble les deux dernières colonnes (traces écrites et traces numériques) parce qu'elles posent des questions voisines – même si les traces numériques demandent en général *plus* d'efforts que les autres pour construire des données offrant un grand potentiel interprétatif. En contrepartie, évidemment, elles permettent d'obtenir plus facilement et rapidement un gros volume de données.

Tableau 1 : Forces et faiblesses des méthodes de collecte de données de réseaux

Critères\Méthodes	Questionnaire (A.)	Entretien (B.)	Expérimentation (C.)	Observation (D.)	Traces écrites (E.)	Traces numériques (F.)
Contrôle des conditions de production (1., 2., 3.)	**	***	***	**	**	*
Statut clair des liens manquants (4.)	***	*	*	*	*	*
Prise en compte des liens négatifs, latents, jetables (5.)	*	**	*	*	**	*
Prise en compte d'éléments sur la substance des relations (6.)	*	***	*	***	**	*
Prise en compte d'éléments sur les attributs des entités (7.)	***	**	**	*	**	*

Légende : plus il y a d'étoiles, plus la méthode de collecte satisfait aisément, dans notre expérience, ce critère. Les numéros et lettres renvoient aux différentes sections de cet article dans lesquelles chacun de ces éléments est traité.

Nous proposons dans ce texte un guide pratique pour initier une recherche mobilisant l'ARS, illustré par une sélection de cas tirés de recherches anciennes ou récentes : des cas que nous connaissons bien, auxquels nous nous référons souvent lors de formations. Ils sont, du fait de nos propres spécialités, choisis en sociologie et en histoire, mais nous pensons, pour les avoir régulièrement présentés à des publics d'autres disciplines, que notre propos s'applique plus généralement aux sciences humaines et sociales.

Écueils et arbitrages dans la construction des données de réseaux

Dans cette partie, nous évoquons les principales questions qu'il est, selon notre expérience, important de se poser lors de la construction des données de réseaux – quelles que soient leurs sources. Ces questions amènent à chaque fois des arbitrages entre aller plus vite et construire des données plus riches. Nous pointons les principaux problèmes que peut poser le fait de choisir systématiquement la voie la plus rapide : ces écueils nous sont régulièrement apparus lors de formations, alors qu'il était souvent trop tard pour refaire la construction des données. Certains peuvent conduire à une sous-interprétation : les données ont été tellement abstraites, séparées de leur contexte, simplifiées, qu'une fois soumises à l'ARS, elles ne permettent pas d'obtenir des résultats qui répondent aux questions de recherche. Dans d'autres cas, c'est plutôt la surinterprétation qui pose problème : on suppose que les données représentent bien tel ou tel type de relation, alors qu'il y a un fort doute sur le fait que ce soit le cas. En effet, les données présentent des biais, au sens où elles ne sont pas adaptées à l'interprétation construite, et ces

biais n'ont pas été vus lors de leur construction. Envisager assez tôt les sept points que nous listons ici peut aider à éviter ces deux écueils très généraux.

Toutes les données ne sont pas bonnes à prendre (1.)

Les données dites « relationnelles » sont aujourd'hui très accessibles. En particulier, le développement des médias sociaux et l'accès aux API (*Application programming interface*) de ces plates-formes rendent le rassemblement de ce type de données relativement facile. Même si Facebook limite fortement l'accès à ses données, d'autres plates-formes comme Twitter ou Youtube laissent plus de marge de manœuvre aux chercheurs·ses en leur permettant d'obtenir (avec des limites de volume néanmoins) des listes de tweets, retweets, likes ou autres commentaires de vidéos. Il est possible d'explorer des problématiques déjà étudiées en sciences sociales avec ces données, mais aussi et peut-être surtout d'en définir de nouvelles. Dominique Boullier (2015) propose par exemple d'étudier ce qu'il appelle les « vibrations », c'est-à-dire l'évolution des entités circulant sur ces plates-formes (évolution d'un message, d'un *hashtag* ou d'un concept). Alors que les « anciennes » sources (données administratives ou issues de sondages) permettraient d'étudier les structures sociales de longue durée ou les mouvements d'opinion, ces nouvelles sources issues des plates-formes permettraient de suivre l'évolution de ces entités en ligne, leur diffusion mais aussi leur réappropriation ou leur transformation au cours du temps et dans différents espaces sociaux. Néanmoins, du fait que ces données peuvent être assez facilement obtenues en très grande quantité, la tentation est forte de s'en saisir pour répondre à tous types de questions – voire pour en faire des graphes sans véritable problématique de recherche préalable.

Ce problème n'est en fait pas nouveau ni spécifique aux traces numériques. Il se pose par exemple en histoire dans les cas où il reste peu de traces écrites d'une société. Ainsi, dans l'Europe de la fin du Moyen-Âge, les seules traces que l'on repère de beaucoup d'individus et de leurs liens sociaux apparaissent dans des actes notariés (contrats de mariage, de ventes de terres, actes de succession). La satisfaction à y repérer des indications de parenté, de voisinage ou d'opérations économiques peut facilement faire oublier que l'immense majorité des liens sociaux pertinents pour ces individus n'ont pas laissé de trace de ce type. Ce n'est pas que les actes notariés ne soient pas bons à prendre pour l'ARS ; simplement qu'il faut se souvenir qu'ils ont été au départ produits par des officiels pour protéger des droits et identifier des populations : leur but n'était pas de documenter toutes les relations sociales, ni même toutes les relations familiales et économiques. De même que le but de Twitter n'est pas de créer une base de données sur la transmission d'opinions.

Tout cela n'interdit pas, bien sûr, d'utiliser tous ces matériaux qui sont relativement plus accessibles que d'autres pour construire des données – y compris lorsqu'on a des questions de recherche très éloignées des motivations des producteurs de ces données. On emploie parfois en histoire l'expression « lire les sources contre leur grain » pour décrire cette opération de détournement volontaire. Mais il est important de se souvenir de ces raisons de production des documents pour ne pas oublier que bien des événements et transactions ne laissent pas de traces écrites ou numériques, pour diverses raisons propres à chaque contexte social : minorités religieuses si l'enregistrement est fait par des prêtres ; petites transactions s'il est coûteux de les enregistrer ; contacts en face à face ou au téléphone par opposition aux lettres ou tweets, etc.

Rappelons une évidence : la problématique est supposée conduire au choix de la méthode de construction des données, et non l'inverse. En sciences humaines et sociales, l'idée n'est pas de tomber sur des données que l'on pense être « relationnelles », de leur appliquer un logiciel de réseau qui produit un graphe, puis de réfléchir à la problématique scientifique qui pourrait s'y appliquer. Les personnes que nous avons rencontrées en formation et qui se retrouvaient dans cette situation ne parvenaient en général pas à construire *a posteriori* une telle

problématique, notamment parce qu'elles n'en savaient pas assez sur ces données trouvées : leur facilité d'obtention s'était retournée en pauvreté menant à une sous- ou surinterprétation. Bien sûr, la disponibilité ou pas de sources ou les difficultés rencontrées pour construire les données peuvent amener à modifier la problématique initiale, mais il est risqué de partir sans elle. L'ARS n'échappe pas à la nécessité de trouver une manière de construire des données adaptées à une problématique.

En revanche, la facilité à obtenir des données peut elle-même faire l'objet d'une réflexion féconde. Se poser des questions sur ces conditions même d'obtention est un premier moyen d'enrichir ses données, de se prémunir contre leur surinterprétation, mais aussi de construire des pistes interprétatives originales. Comment se fait-il que je sois assez facilement « tombé·e » sur ces données ? Qui les a construites et qui les a rendues disponibles ? À qui ? Dans quel but ? Quelle était la problématique de ces personnes ? Est-il possible d'utiliser les données qu'elles ont produites pour répondre à des questions différentes des leurs ? Et, question symétrique, souvent plus importante encore : qu'est-ce qui *n'est pas* dans ces données ? Quelles entités, quels liens, quelles précisions sur les liens sont absentes ? Il est parfois possible de s'informer directement (par tout mode d'enquête, de l'entretien aux traces numériques) sur ces conditions de collecte des sources et de construction des données. D'autres fois, ce n'est pas possible : pensons à une inscription antique que l'on est déjà bien heureux·se de trouver, et qui n'est pas accompagnée d'archives sur la manière dont son texte a été décidé ni sur le salaire du tailleur de pierre... Toutefois, l'observation fine et/ou quantifiée des données elles-mêmes, dès lors que l'on a en tête les questions ci-dessus, peut permettre de trouver des éléments de réponse sur les conditions de leur production, et donc sur ce qu'elles peuvent et ne peuvent pas permettre de montrer. Toutes les sources peuvent être bonnes à prendre pour les sciences humaines et sociales, y compris pour l'ARS ; mais des données dont on ignore le mode de construction sont à considérer avec prudence.

Les données ne sont qu'un point de vue sur un lien (2.)

On n'observe jamais directement un ensemble de liens sociaux : même dans le cas d'un travail ethnographique, notre regard est limité dans le temps et l'espace et tributaire d'une définition préalable, plus ou moins explicite, de ce qui nous intéresse. Si l'on part, par exemple, avec l'idée très générale « je vais aller dans une classe et noter qui est ami·e avec qui, en observant plutôt qu'en demandant aux enfants », il faut tout de suite commencer à se demander quels éléments d'observation, concrètement, pourraient conduire à noter « ami·e » – ou plutôt pourraient, dans une phase ultérieure de recherche, être résumés ainsi. Et se demander ce que l'on va rater (quels indices, quelles amitiés) en étant dans une certaine position, à un certain moment, dans la classe, et en étant soi-même, avec ses caractéristiques propres.

C'est ce type de définition, en général affiné au fil d'une recherche, qui nous amène à considérer certains liens comme similaires et à les compter ensemble, à en catégoriser d'autres séparément, et à ne pas en noter une grande quantité. La même chose est vraie *a fortiori* lorsqu'on travaille par entretien, questionnaire, sur des documents produits sur notre terrain d'étude ou des données mises en forme par d'autres chercheur·ses. Non seulement les données ne sont pas « données » mais construites à partir de sources (c'est-à-dire des matériaux plus compliqués, moins structurés que nous recueillons), mais les sources ne sont pas toute la réalité sociale. C'est un truisme des sciences humaines et sociales, mais nous voulons insister ici sur la nécessité de le garder en tête pendant tout le travail de recherche, jusqu'au moment de la production et de l'interprétation de résultats.

Cela implique notamment qu'un tableau de liens, une matrice ou un graphe doit être assorti d'explications sur les catégories exactes de liens et d'entités qui y sont figurées et qui en sont exclues, et qu'il faut indiquer qui a défini ces catégories. En effet, selon que notre source

émane de tel ou tel acteur social, de telle ou telle institution, et selon les choix de construction de données que nous avons nous-mêmes faits, très concrètement, les résultats ne seront pas les mêmes : ce n'est pas le même acteur qui va être le plus central dans « le réseau », ou encore les liens apparaîtront réciproques, ou non.

Par exemple, dans les mairies de France, au XIX^e siècle, chaque époux·se devait être accompagné·e de deux témoins au mariage (pas zéro, pas cinq – et deux hommes) ; le plus souvent, le lien du témoin à l'époux·se est précisé dans l'acte de mariage lorsque ce lien est familial (beau-frère, oncle, etc.) et décrit comme « ami » ou non décrit dans tous les autres cas. Ces derniers cas incluent des personnes voisines de la mairie, appelées à jouer régulièrement le rôle de témoins pour les époux·ses qui ne sont pas venu·es avec les leurs (Gourdon, 2008). Au contraire, dans les contrats de mariage notariés, qui ne sont établis que pour les personnes qui ont au moins un petit patrimoine, le nombre de témoins est au choix des parties. Étudier, par exemple, les liens de témoignage au mariage entre différents métiers ou différents lieux donne dès lors des résultats très différents selon qu'on se fonde sur les actes ou sur les contrats de mariage. *A fortiori*, étudier qui se dit « ami·e » de qui dans des correspondances ou des journaux intimes, par exemple, ou bien dans des actes de mariage donnerait des résultats extrêmement différents ; de même pour une collecte d'« ami·es » Facebook (une catégorie fixée par la plateforme et nécessairement réciproque, mais dont les usages et contenus vécus diffèrent énormément selon les personnes) ou encore une série d'entretiens demandant simplement « qui sont vos ami·es ? », ou bien proposant aux enquêté·es une définition plus précise de ce type de relations. Ainsi, James Kitts et Diego Leal (2021) comparent les définitions de l'amitié recueillies dans des entretiens collectifs auprès de collégien·nes à celles qu'envisagent généralement les auteur·es de questionnaires. Aucune de ces sources n'est en soi meilleure que l'autre, ou plus proche de « l'amitié » dans l'absolu ; mais elles présentent des biais différents pour étudier ce phénomène.

Lorsqu'on construit des données pour l'ARS, il faut donc se livrer aux mêmes opérations de réflexivité (comme on dit en sociologie) ou de critique des sources (comme on dit en histoire) que lors de tout autre type de recherche (voir par exemple Cefai, 2003 ; Paugam, 2012 ; Boutier 2001 ; Ruiz 2011). En particulier, il faut comprendre (et dire, ensuite, lorsqu'on publie) ce qui est très probablement exclu ou sous-représenté dans les données, ce qui fait qu'un lien va être compté comme existant, qui (acteurs sociaux étudiés et/ou analystes) l'a défini et comment. Des sociologues et criminologues ont très bien discuté ce problème dans une revue de littérature portant sur l'analyse des réseaux criminels à partir de sources policières et judiciaires (Bright et alii, 2021). Comprendre ce que la source tait est nécessaire pour interpréter les données : il s'agit notamment de comprendre quels aspects des graphes ou des chiffres découlent mécaniquement des biais de la source (de son point de vue particulier) et lesquels nous disent quelque chose d'autre de la réalité étudiée. Dans des cas extrêmes, la valeur de certaines analyses de réseaux réside avant tout dans ce qu'elles permettent de mieux comprendre de la structure de leur source, ce qu'elle dit et promet et ce qu'elle cache, plutôt que de la structure du monde social ; c'est le cas de l'étude par l'historienne Mathilde Jourdan (2019) des rencontres et des déplacements relatés dans des vies de saints du VII^e siècle.

On n'observe qu'un pan de la réalité (3.)

Parmi les limites de l'observation liées à la fois aux possibilités d'accès aux sources et aux choix de recherche, il y a le choix de l'échantillon ou du groupe étudié : on ne peut jamais étudier tout le monde (toutes les personnes ou autres entités qui nous intéresseraient dans l'absolu), ni tous les types de liens. Dans la réalité, il est bien rare que les liens étudiés n'existent qu'au sein de l'échantillon ou du groupe de personnes (ou autres entités) qu'il est possible d'étudier. Il faut donc dire clairement quel est le périmètre de l'étude et ce que cela implique

pour ses résultats. Par exemple, une personne peut apparaître isolée dans un graphe ou selon un indicateur non pas parce qu'elle n'est pas impliquée, en général, dans le type de lien étudié, mais parce qu'elle n'a pas de lien avec les autres personnes étudiées (seulement avec des personnes non étudiées) – ou parce que son lien n'apparaît pas au moment étudié, ou dans la source prise en compte (il n'a pas été enregistré, ou n'est pas connu de tout le monde). Par ailleurs, en aucun cas les frontières des réseaux sociaux ne correspondent « naturellement » à celles de systèmes sociaux étudiés : ces frontières sont construites par les chercheurs·ses pour répondre à une question spécifique et il ne peut y avoir de démarche scientifique sans ces formes de découpage (Laumann et alii, 1983).

En général, les personnes qui participent à des formations en ARS sont très préoccupées par la volonté de rassembler la plus grande quantité possible de données (de documenter beaucoup de liens). Elles sont dès lors souvent surprises que nous insistions beaucoup sur les opérations d'abstraction ou de découpage. Pourtant, comme pour toutes recherches, ces opérations sont essentielles. Cela vaut en aval, tout particulièrement pour les graphes d'ARS, qui deviennent vite difficiles à lire : leur mise en lisibilité passe souvent par le fait de *caler* provisoirement une partie de l'information. Mais cela vaut aussi en amont : dès lors qu'on n'observe jamais *tous* les liens existants d'un certain type, mieux vaut un découpage clair, c'est-à-dire savoir ce que l'on rate. Cela implique tout d'abord de se poser la question de l'échelle d'observation adaptée à la question posée.

Ainsi, les données de réseaux construites par extraction depuis les plates-formes numériques sont souvent bien plus massives que celles construites par d'autres méthodes (questionnaires, archives ou observations). Bien souvent, ces réseaux présentent aussi des structures très modulaires, c'est-à-dire qu'ils sont composés de sous-graphes relativement denses et relativement déconnectés les uns des autres, lorsqu'on les regarde dans leur globalité. Un réseau de réponses et retweets (liens entre comptes) extrait sur la base d'un *hashtag* particulier, par exemple, présente souvent une structure polarisée entre plusieurs sous-graphes denses, qui peuvent correspondre à une sensibilité politique : dans chaque sous-graphe, les différents comptes se renvoient les uns les autres, mais très peu entre sous-graphes. Les algorithmes de détection de communautés – par exemple la méthode dite « de Louvain » (Blondel et al., 2008) – s'appliquent parfaitement à ce type de données (comme aux réseaux de citations entre articles scientifiques extraits automatiquement du Web of Science, par exemple) ; comme la méthode est accessible en mode « pousse-bouton » dans certains logiciels, il est devenu courant de voir des cartographies de « communautés » présentées sans problématisation explicite (Stoltenberg et al., 2019). Il ne s'agit pourtant que d'une des échelles d'analyse possibles, y compris pour ces données. Si l'on analyse plus spécifiquement chacun des sous-graphes, on rencontre souvent des structures différentes, qui peuvent se rapprocher de celles des « petits » réseaux de quelques dizaines d'individus collectés par questionnaires sociométriques (voir ci-dessous). Ceux-ci ont souvent une structure plus cohésive et hiérarchique, avec un centre et des périphéries, et plus complexe, avec des relations transitives, par exemple – ce que les algorithmes de détection de communautés ne mesurent pas. Si l'on veut visualiser l'ensemble des données simultanément dans les grands réseaux issus de plates-formes, on choisit donc une échelle parmi plusieurs possibles : une vue d'en haut qui implique souvent une assez mauvaise connaissance des dynamiques locales. La définition de l'échelle pertinente d'observation du réseau dépend donc de la question de recherche – comme celle du critère à employer pour délimiter les frontières des données.

Quelle que soit l'échelle, il y a trois approches principales pour délimiter cette partie de la réalité à observer : en partant d'un groupe, d'une personne ou d'un type de liens. Chacune de ces approches peut *a priori* s'appliquer à tous les types de sources que nous évoquerons en

seconde partie³. Par ailleurs, ces différentes approches dépendent tout autant de choix et contraintes méthodologiques que de choix théoriques.

Premièrement, on peut définir l'échantillon ou le groupe à étudier par un attribut que partagent les entités en cause (présenter telle caractéristique, appartenir à telle organisation, participer à tel événement, etc.) et se donner pour but de reconstituer ensuite la présence ou l'absence de relations entre *toutes* les dyades possibles de ces entités (toutes les combinaisons possibles de deux de ces entités). C'est ce qu'on appelle parfois l'approche « sociocentrée » et souvent l'approche en réseaux « complets ». Cette appellation se prête aux malentendus, toutefois : il ne s'agit pas de rechercher *tous les types* de liens entre ces individus, mais seulement un ou quelques types de liens précis ; en outre, le réseau construit n'est « complet » qu'au sein de l'échantillon ou du groupe défini au départ. Dans la réalité, il arrive souvent que les entités que nous étudions entretiennent aussi les liens qui nous intéressent à l'extérieur de ce groupe. Le réseau est donc « complet » pour le lien choisi, au sein de l'échantillon et du groupe choisis. Cette approche a par exemple été appliquée pour documenter par questionnaire toutes les relations d'échanges de conseils d'ordre professionnel entre les membres d'une même organisation, ou, sur la base de sources écrites, tous les partages d'administrateurs au sein d'un ensemble précis de grandes entreprises.

Deuxièmement, on peut plutôt se centrer sur une série d'individus appelés *egos*, qu'on choisit suivant un critère lié à l'étude envisagée. On étudie alors tous leurs liens d'un certain type. Ces relations nous amènent à des *alters* : les individus liés à *ego*. On essaie ensuite de documenter l'ensemble des liens entre les *alters* (ce qui peut donner des résultats différents selon que l'existence de ces liens est affirmée par les *egos*, par les *alters* ou par des tierces personnes). C'est ce qu'on appelle une approche « égocentrée ». Elle a été appliquée à des études sur l'amitié aussi bien que, par exemple, à celle de citations entre des ensembles d'articles scientifiques. Par définition, tout le monde aura un lien avec un des *egos* : comme pour les deux autres approches, le mode de construction des données joue sur la structure obtenue. C'est pourquoi il est important de le garder en tête et, autant que possible, de le choisir en fonction de la question de recherche plutôt que d'aller au plus simple ou au plus rapide.

Troisièmement – ce dernier cas n'est en général pas listé dans les manuels, mais on le rencontre souvent en pratique –, il arrive que l'on parte d'une liste de liens. C'est une série d'interactions ou de relations que l'on observe, que l'on trouve dans la source écrite ou les traces numériques (par exemple, toutes les transactions enregistrées par un notaire pendant une année ; tous les *tweets* d'une journée qui emploient un *hashtag*), ou c'est à propos de relations et d'interactions (plutôt que de noms de personnes) que l'on pose des questions en entretien ou par questionnaire, ou ce sont des liens que l'on produit dans le cadre d'une expérience (de reconstitution de « chaînes relationnelles » : voir ci-dessous). On ne part donc pas d'une liste d'individus pour rechercher leurs liens : ce qui vient d'abord, c'est la liste de liens. Dans ce cas, toutes les entités apparaîtront nécessairement au sein d'un lien : aucune ne sera isolée, alors que cela peut arriver avec les deux autres approches. Cette troisième approche pour délimiter l'objet d'étude n'a pas de nom dans la littérature ; on pourrait l'appeler « lien-centrée ».

Dans les trois cas, il faut enfin se demander si les critères employés pour définir si les entités sont incluses ou non, et si un lien est compté comme existant ou non, sont « réalistes » ou « nominalistes » (selon le vocabulaire de Laumann et alii, 1983 ; on pourrait aussi dire *emic* ou *etic*). Les critères sont « réalistes » si l'analyste tente de reproduire une définition qui fait sens sur le terrain enquêté (il faut alors se demander qui emploie cette définition, si elle est contestée ou non, etc.). Un *hashtag* partagé est typiquement un critère réaliste, particulièrement facile à mettre en œuvre (chaque tweet le contient ou pas) mais, en contrepartie, difficile à

³ Notre présentation des deux premières approches s'inspire de Laumann et alii, 1983 et Eloire et alii, 2011, qui donnent plus de détails, dans l'optique d'enquêtes par questionnaires ou entretiens.

interpréter. Les critères sont « nominalistes » si l'analyste propose sa propre manière de catégoriser les attributs des personnes ou des liens et se fonde sur ses critères, qui doivent être explicites, pour définir le périmètre de l'enquête – par exemple s'il ou elle s'emploie à étudier les retweets au sein d'une Twittosphère féministe qu'il ou elle a définie au préalable.

Les liens manquants comptent (4.)

« L'ARS n'est pas évidente à importer de la sociologie vers X, parce qu'en X, on a beaucoup de données manquantes. » Remplacez X par votre discipline et vous retrouverez peut-être le scrupule le plus entendu en formations... En réalité, la sociologie, quelle que soit la méthode de collecte des données employée, fait aussi face à ce problème des données manquantes. C'est ce que nous disons pour rassurer les personnes formées : on peut quand même faire de la recherche quand on n'a pas une information complète sur tout... Et on a bien vu dans les points précédents que, quelle que soit l'approche choisie pour délimiter l'objet de l'enquête, l'ARS ne peut en général porter que sur une petite partie de l'ensemble des liens qu'il serait en théorie pertinent d'étudier. Mais il reste vrai que la question des données manquantes se pose de manière plus aiguë que pour d'autres approches en sciences humaines et sociales. En effet, la forme d'un réseau est donnée par les liens absents, tout autant que par les liens présents. L'existence de « petits mondes » et d'« intermédiaires », notamment, est déduite de l'absence de certains liens : il y a plus de liens au sein du petit monde qu'entre lui et l'extérieur ; l'intermédiaire a une position particulière parce qu'il relie deux groupes entre lesquels les autres liens potentiels sont absents. Les liens absents définissent des positions de pouvoir (du fait d'un monopole sur certains liens) ou d'isolement, ou encore des frontières.

Nous avons vu beaucoup de graphes produits par des débutant·es qui essayaient ainsi d'interpréter une position d'intermédiaire comme une position de pouvoir, alors même qu'elle découlait mécaniquement d'un simple manque d'informations sur certains liens. C'est le cas typiquement dans les réseaux de correspondances (qui écrit à qui), qui ne sont en général connus qu'en partie : ce sont des collections de réseaux égocentrés (chaque récepteur de lettres pour qui une collection existe est un *ego*). Y interpréter l'intermédialité comme s'il s'agissait d'un réseau sociocentré revient à confondre un « je ne sais pas » (je ne sais pas si Albert écrivait à Mathilde, car il n'existe pas de collection de lettres reçues par Mathilde) et un « non » (je suis sûr·e qu'Albert n'écrivait pas à Mathilde). Les visualisations habituelles de réseaux, pour lesquelles seuls les liens avérés sont représentés par des traits, tandis que les absences de lien n'ont pas de représentation propre, peuvent inciter à négliger cette question redoutable : peut-on distinguer une absence d'information d'une absence de lien ?

Comme pour les points précédents, l'arbitrage consiste donc ici à allouer du temps, ou non, à réfléchir à ce qui est *absent* des données et à la manière dont cette absence peut peser sur leur interprétation. Il s'agit d'explorer toutes les pistes permettant de distinguer les non-liens probables des inconnues complètes. Dans un des rares articles discutant explicitement du sens des « zéros », Alexis Ferrand (2014) rappelle que les méthodes classiques (dites de générateurs de noms – voir plus loin) demandant d'énoncer des listes de noms ne donnent pas l'exhaustivité du réseau relationnel : par exemple, dans la réalité, si la première personne à laquelle on penserait spontanément à demander de l'aide n'est pas disponible, on fait souvent appel à une autre. Le fait de ne pas lister une relation en réponse à une question ne signifie pas qu'elle n'a aucune probabilité d'exister à un moment, mais il est très difficile d'enquêter sur cette probabilité. Parfois, il est possible de mieux comprendre et interpréter les liens manquants (plutôt que de les interpréter comme inexistant) en comparant deux sources de données, comme le fait la médiéviste Isabelle Rosé (2018), montrant que les narrations et les traces d'actes officiels donnent deux images différentes des relations autour de la reine Emma. D'autres fois, recourir à plusieurs hypothèses de reconstitution différentes et comparer les

résultats peut être le moyen de ne pas considérer d'emblée une information manquante comme une absence de liens.

Les liens ne sont pas forcément des ressources (5.)

Trop souvent, les relations sont considérées seulement ou prioritairement comme des ressources ou comme un moyen d'accéder à des ressources. La métaphore du capital social (chez Bourdieu, 1980 ; Putnam, 2000 ; Coleman, 1988 ; Burt, 2000 ou Lin, 2002) sur laquelle s'appuient nombre d'analyses des relations y contribue beaucoup. Cette notion est bien entendu heuristique pour expliquer les inégalités, les parcours sociaux ou encore des formes de pouvoir. Mais, en limitant les relations ou les structures de relations à un stock qu'il serait possible de réinvestir, elle conduit les chercheurs à se focaliser sur les relations potentiellement les plus bénéfiques pour un individu, en mettant de côté beaucoup d'autres relations qui influencent tout autant les parcours sociaux.

Des travaux récents ont ainsi réfléchi à la manière de collecter et étudier des liens non seulement perdus ou oubliés – un type bien particulier de données manquantes – mais aussi négatifs (Chauvac et alii, 2014 ; Roda, 2014 ; Harrigan et alii, 2020). Offer et Fischer (2018) appellent liens « pesants » ou « difficiles » (*burden ties*) ces relations que l'on maintient malgré des rapports interpersonnels tendus : un collègue avec qui l'on est en dispute, un parent âgé et demandeur, un enfant avec qui les relations sont conflictuelles. Ces relations inévitables influencent considérablement les parcours sociaux et restent pourtant souvent en dehors du radar des méthodes utilisées pour étudier les réseaux personnels, si l'on n'y dédie pas des questions spécifiques pour les mesurer. On peut ajouter, pour des cas de liens autres qu'entre personnes physiques, une délégation qui s'oppose à une autre dans une assemblée, un article qui en cite un autre pour le critiquer, etc. Ces liens apparaissent rarement dans les articles d'ARS, alors qu'ils étaient analysés dans des travaux pionniers sur des petits groupes (voir par exemple White et alii, 1976). Pourtant, rien n'empêche d'enquêter à leur sujet, pour comprendre leur genèse (Roda, 2014) ou parce qu'ils participent aussi à structurer les autres relations : ainsi, les relations d'amitié ou d'alliance se forment très souvent par des relations d'inimitiés communes (Pál et alii, 2016 ; pour un bilan et un programme de recherche plus généraux sur les liens négatifs, Harrigan et alii, 2020). Mais il faut tenir compte du fait que les indicateurs classiques ont été plus ou moins implicitement construits pour traiter de liens positifs. Non seulement l'interprétation des liens qui ne sont pas des ressources est particulière, mais la production d'indicateurs chiffrés à leur sujet, et même leur visualisation, obligent à sortir des procédures les plus habituelles de recherche. Ici encore, on a donc un arbitrage : en rester aux liens-ressources est en général moins coûteux en temps et en inventivité, mais enrichir les données par d'autres liens peut être pertinent par rapport à la problématique choisie.

Il est aussi important d'avoir à l'esprit que les liens qui procurent effectivement des ressources ne sont pas forcément les liens qui sont cités ou considérés comme procurant ces ressources. Par exemple, les générateurs de noms, fréquemment utilisés dans les études de réseaux personnels, demandant de lister les personnes dont on se sent proche ou celles avec qui l'on parle de problèmes importants, font remonter des personnes qui ne sont pas forcément celles de qui l'on obtient du support social. En règle générale, les répondants obtiennent ce support de personnes accessibles, disponibles, et plus à même de comprendre le contexte du problème, et non pas de proches (Small, 2013). Dans certains cas, comme celui de la grande pauvreté aux États-Unis, ce sont même des liens considérés comme « jetables » (*disposable*) entre personnes qui ne se connaissent pas, c'est-à-dire des personnes avec qui la construction d'une intimité est accélérée et/ou simulée, qui procurent certaines ressources vitales dans des cas critiques (Desmond, 2012).

Ce cas des liens « jetables » pose la question de la temporalité de ce que l'on étudie. D'une façon générale, en ARS, il est particulièrement utile de noter toute date disponible (même approximative) et de se poser la question de la durée des liens. S'agit-il de relations auxquelles on peut attribuer une date de début et – question plus difficile, mais souvent importante – de fin, ou d'affaiblissement ? Ou bien d'interactions plus ponctuelles, mais peut-être répétées ? La réponse à ces questions orientera le traitement visuel ou quantitatif des données (Lemerrier, 2015), mais peut aussi, en amont, faire privilégier un type de source ou un autre, selon qu'il permet ou non une observation continue sur une durée un peu longue, ou au contraire plus ponctuelle mais peut-être plus détaillée ou exhaustive.

Documenter la substance des relations permet de mieux interpréter leur forme (6.)

L'ARS peut apparaître comme une méthode visant à l'abstraction, à la montée en généralité. Elle propose en effet un vocabulaire qui vise à désigner des formes, des structures de relations similaires qui se retrouvent dans des situations très différentes du monde social. On peut par exemple avoir une position « d'intermédiaire » entre deux groupes par ailleurs relativement isolés l'un de l'autre lorsqu'on est la seule personne qui connaît l'identité des membres de deux groupes impliqués dans la résistance, aussi bien que lorsqu'on est la seule entreprise qui fait de l'import-export entre deux pays. Dans les deux cas, cette position peut être à certains égards qualifiée de position de pouvoir, notamment parce qu'elle peut donner un certain contrôle sur des flux d'informations. On peut aussi la voir comme un point de vulnérabilité (si l'intermédiaire disparaît, les deux groupes sont isolés), et dans le premier cas comme une cible privilégiée de la répression. Mais on ne peut pas aller très loin dans l'interprétation en considérant une position d'intermédiation dans l'absolu, sans prise en compte du type de lien concerné (Crossley, 2010). Pourtant, dans certains domaines de recherche, les indicateurs de réseaux sont réifiés comme s'ils étaient des réalités en soi. C'est le cas dans une partie des sciences de gestion étudiant des réseaux inter-organisationnels, par exemple, domaine où les réseaux étudiés sont souvent composés de contrats, partenariats ou *joint-ventures* entre entreprises. Nombre d'articles dans ce domaine présentent la centralité de degré (c'est-à-dire le nombre d'arêtes adjacentes d'un nœud dans un graphe, ou plus simplement le nombre de relations d'une personne dans un réseau) comme un indicateur de pouvoir dans un système d'échanges ; l'intermédiation serait liée, elle, à la performance économique d'une entreprise, et la faible densité à un manque d'opportunités économiques (par exemple Boje et Whetten, 1981 ; Williams, 2005 ; Shipilov et Gawer, 2020). Ces indicateurs, tout comme les visualisations de graphes, ne font pourtant sens qu'en tenant compte du type de lien étudié. Le fait de trop souvent considérer les relations comme des ressources est probablement l'une des causes de cette réification des indicateurs. Pour prendre un exemple trivial, être central dans un réseau de discussions dans une organisation n'est pas équivalent à être central dans un réseau d'inimitié.

En formations, nous insistons donc souvent sur le fait qu'un graphe représentant « Le réseau de X », sans que le titre ni la légende ne rappellent quels sont les liens pris en compte et d'après quelles sources, pose problème – tout en avouant qu'il est difficile de produire un titre à la fois court, percutant et juste, bien sûr. Mais dans l'idéal, il faudrait que le lectorat ou l'auditoire se souvienne toujours de quel type de lien on parle. En général, la problématique de recherche est en effet particulière à une forme d'interaction, d'échange ou de relation. Amitié, amour, collégialité, parenté ne sont pas assimilables de manière évidente à des relations économiques, de pouvoir ou d'influence. Un contrat de fourniture entre entreprises ne fonctionne *a priori* pas comme une recommandation entre personnes ou une citation entre articles : si l'on pense que c'est pourtant le cas, cela mérite d'explicitier pourquoi. Et même si le but est d'atteindre, à un moment, ce niveau d'abstraction, il est important, lors de la construction des données, de rester très clair·e tant sur le type d'entités concernées que sur les

autres caractéristiques de la relation : comment elle est nommée par différents acteurs, comment on peut éventuellement attester sa réciprocité, son intensité, sa durée, etc. Bref, ne pas s'en tenir à « ce lien existe-t-il, oui ou non ? » mais le caractériser par le plus grand nombre d'attributs possibles. Cela vaut pour les liens considérés comme des ressources mais aussi pour les autres : ainsi, une étude récente distingue (dans le cas de données construites par questionnaire sur des relations entre lycéen·nes) évitement, antipathie et agression (Kros et al., 2021).

Ensuite, les analyses ne viseront pas à additionner ensemble tous ces liens bien caractérisés (et encore moins à les représenter tous en même temps sur un graphe multicolore et compliqué). En général, il est en effet plus intéressant de procéder par comparaison que par accumulation. Par exemple, l'article classique de John Padgett et Christopher Ansell (1993) sur l'arrivée au pouvoir des Médicis à Florence à la Renaissance étudie à la fois des relations d'alliance et des relations économiques entre les familles de l'élite locale. La position des Médicis dans ce réseau multiplexe (composé de deux grands types de liens) est particulière parce que leurs liens économiques ne concernent pas les mêmes familles que leurs liens d'alliance. Additionner les deux types de liens, plutôt que de les comparer, n'aurait pas permis de repérer cette position.

Collecter des données sur les entités pour mieux comprendre leurs relations (7.)

Un autre point essentiel pour éviter de se retrouver en situation de sous-interprétation des résultats de l'ARS est de disposer d'informations sur les entités prises en compte (individus, organisations, lieux, etc.), en plus de leurs seuls noms et liens. Ces autres informations, appelées traditionnellement « attributs » en ARS, par différence avec les « liens », sont aussi des données construites par le chercheur ou la chercheuse.

C'est là l'un des principaux écueils des matériaux issus des plates-formes numériques. Si la plupart des plates-formes collectent des données sur leurs usagers, les chercheurs·ses n'ont souvent accès qu'aux noms des comptes et à leurs nombres d'abonnés, et dans de très rares cas au genre, à l'âge, à la localisation, au métier, etc. des personnes qui tiennent ces comptes. Lorsqu'on ne dispose d'aucune information de ce type, on doit se contenter d'une analyse formelle centrée sur des nœuds et des arêtes, considérés comme des objets abstraits, telle qu'elle est pratiquée par les théoricien·nes des graphes. Or, les nœuds que nous étudions sont des entités, souvent des acteurs sociaux (et économiques, scientifiques, etc.) dont les caractéristiques sociales peuvent expliquer leurs relations, et inversement. En sciences humaines et sociales, les données relationnelles ont peu d'intérêt si elles ne sont pas associées à des données décrivant certaines propriétés sociales des entités.

Un risque de sur-interprétation existe aussi si on imagine, faute de telles informations, que les entités qu'on observe sont représentatives d'une population plus large, alors que ce n'est pas le cas. Ainsi, les usagers des plates-formes comme Twitter sont plus diplômés et plus favorisés que la moyenne et certaines professions sont clairement sur-représentées (journalisme, communication, recherche) – comme le montrent les travaux qui utilisent, en complément de l'extraction de données numériques, d'autres méthodes pour décrire cette population (Boyadjian, 2014). Il faudrait en tenir compte dans l'interprétation des données qui en sont issues, alors même qu'on ne peut pas précisément mesurer ce biais. Le propos de l'ARS n'est en effet pas d'oublier les catégories autres que relationnelles de la sociologie (classes, communautés, groupes, nations, etc.), mais au contraire de les étudier en même temps que les relations et interactions (amitié, échange d'informations, lien familial, co-présence dans une entreprise, citation, etc.) – sans systématiquement survaloriser ces dernières (Godechot, 2012).

Enfin, la temporalité des réseaux n'est pas seulement celle des liens mais aussi celle des entités, qui naissent, meurent, rejoignent ou quittent le groupe ou le lieu étudié. Il est important de noter toutes les indications disponibles en termes de dates (et de lieu), non seulement pour

les liens, mais aussi pour les entités et leurs attributs, qui peuvent également changer pendant la période étudiée.

Se lancer dans une ARS implique donc d'effectuer une double collecte de données : sur les liens et sur les attributs des entités qui sont liées. Selon les cas, cette double collecte mobilise les mêmes sources et méthodes ou deux sources et méthodes différentes. Mais les mêmes principes s'appliquent : la réflexivité ou la critique des sources permet de rendre explicite le choix des caractéristiques pertinentes pour la problématique.

La variété des enquêtes et des sources

Nous allons maintenant parcourir les colonnes du Tableau 1 ci-dessus, c'est-à-dire évoquer plus précisément l'une après l'autre les principales méthodes de construction de données de réseaux et la manière dont il est possible d'éviter les écueils :

- par questionnaire : qu'il s'agisse de sondages sur des milliers de personnes ou d'études de petits groupes, il s'agit de la source la plus connue de données pour l'ARS. Ils visent le plus souvent à produire des listes de noms d'individus ou d'organisations, qui constituent en réalité des listes de liens, ainsi que des données complémentaires permettant de qualifier ces interactions ou relations. Dans un second temps, ces réponses sont converties en données quantitatives, après plusieurs opérations de codage plus ou moins fines. Plus les données sont détaillées, plus les interprétations possibles seront riches ;
- par entretien : en général semi-directifs, les entretiens sont centrés sur la recherche de relations, en fonction de la problématique choisie. Ils permettent à l'enquêteur ou enquêtrice d'obtenir plus facilement des informations qualitatives sur les personnes citées et sur leurs relations, formulées selon les termes des enquêté-es et associées à différents lieux, moments et contextes de mobilisation ;
- par expérimentation : les données collectées sont produites par une situation expérimentale artificiellement créée pour l'occasion. Cette manière de faire est plus commune et moins critiquée en psychologie et en économie qu'en sociologie. Cependant, elle a déjà été utilisée avec profit en ARS ;
- par observation : ce mode de collecte de données est plus rare et n'est presque jamais utilisé seul. Il ne peut être utilisé que dans le cas où le chercheur ou la chercheuse connaît déjà bien les enquêté-es ;
- par accès à des traces de relations : la différence avec les situations précédentes est que ces traces n'ont pas été produites par une interaction avec le chercheur ou la chercheuse, mais pré-existent à sa recherche. C'est le cas des objets trouvés dans des fouilles archéologiques et des archives historiques, qui peuvent constituer ou porter des traces de relations très anciennes, mais aussi de bien d'autres types de documents, jusqu'aux traces d'interactions nativement numériques les plus récentes, matière première de ce qu'on appelle souvent le *big data*. Le point commun est que ces données n'ont pas été construites en fonction de notre problématique : nous devons les prendre comme elles sont (nous ne pouvons pas toujours étendre la population étudiée, ou ajouter des attributs sur les entités, par exemple) et nous demander comment elles ont été construites par d'autres. Elles sont, en quelque sorte, de seconde main, ce qui présente quelques avantages et inconvénients particuliers.

Dans chaque cas, nous revenons sur les précautions d'emploi listées en première partie, tout en présentant les grandes lignes de ces méthodes utilisées pour accéder aux sources et construire les données. Ce catalogue nous a semblé nécessaire car les manuels usuels d'ARS

(du moins en français et en anglais) couvrent très inégalement ces méthodes (Tableau 2) : la plupart se focalisent sur les questionnaires.

Tableau 2 : Les méthodes de recueil de données décrites dans les principaux manuels d'ARS en anglais et français

Critères\Méthodes	Questionnaire (A.)	Entretien (B.)	Expérimentation (C.)	Observation (D.)	Traces écrites (E.)	Traces numériques (F.)
<i>Conducting Personal Network Research: A Practical Guide</i> (McCarty et alii, 2019)	p. 50-52, <u>72-93</u>	p. 98-99				p. 98
<i>Egocentric Network Analysis</i> (Perry et alii, 2018, Partie II)	p. <u>37-56</u> , <u>62-128</u>			p. 56-57	p. 57-60	p. 60-62
<i>Sociologie des réseaux sociaux</i> (Mercklé, 2016)	p. 17, 34, 39, 68					
<i>Social Network Analysis for Ego-Nets</i> (Crossley et alii, 2015, chap. 3)	p. <u>44-61</u> , <u>73</u>	p. 61-63		p. 63-65	p. 65-67	<u>p. 67-72</u>
<i>Doing Social Network Research</i> (Robins, 2015)	<u>p. 51-62</u> , <u>92-126</u> , <u>130-139</u> , <u>149-156</u>	p. 126-127	p. 12, 133-134, 225-226	p. 13, 101, 127-128	p. 101, 128, 140, 145, 158-159	p. 101, 129-130, 144-145, 157-158
<i>Réseaux sociaux et structures relationnelles</i> (Lazega, 2014)	<u>p. 15-37</u>				p. 23-24	
<i>The SAGE Handbook of Social Network Analysis</i> (Scott et Carrington, 2011)	p. 21, 106-108, 273, <u>370-403</u> , 545-547	p. 22, 411-412	p. 523-524	p. 410-411	<u>p. 274-281</u> , 315, 412	<u>p. 167-179</u> , 272
<i>Models and Methods in Social Network Analysis</i> (Carrington et alii, 2005, chap. 2)	<u>p. 10-23</u>			p. 22, 25	p. 24-25	p. 25
<i>Les réseaux sociaux</i> (Degenne et Forsé, 2004)	p. 21-27		p. 17-21		p. 180-182	
<i>Social Network Analysis</i> (Scott, 2000)	p. 4, 77-78, 168-169	p. 36, 78-79		p. 17-19	p. 79-80	
<i>Social Network Analysis: Methods and Applications</i> (Wasserman et Faust, 1994)	p. 45-48, 51-53, <u>55-63</u>	p. 48-49	p. 52-54	p. 49-50, 58	p. 50-51, 61-62, 64-66	p. 56, 62-63

Nous listons ici, de manière très inclusive, tous les passages où sont donnés des conseils pour la réalisation de chaque type d'enquête ou qui donnent quelques informations sur la construction de données dans des études classiques. Les passages de 5 pages au moins sont soulignés.

Pour chaque méthode, nous revenons sur les principaux écueils et arbitrages évoqués en première partie ; nous explicitons ce lien entre les deux parties de l'article en faisant référence, entre parenthèses, aux lignes et colonnes du Tableau 1 ci-dessus. Par exemple, nous indiquons

(L6) lorsque nous évoquons des éléments liés au 6^e critère, énoncé comme '6.' à la fois dans le texte et dans le Tableau 1 : « Documenter la substance des relations » ; nous indiquons de la même manière '(A)' pour faire référence à la colonne « Les questionnaires » dans le Tableau 1.

Précisons que ces grands modes de collecte de sources permettant la construction de données relationnelles ne sont pas exclusifs les uns des autres : en pratique, ils sont d'ailleurs souvent utilisés de manière combinée – ce qui permet parfois de faire avec une méthode ce qui serait impossible avec une autre. Pour la clarté de l'exposition, nous les présentons ici tour à tour, en évoquant quelques enquêtes, sans entrer dans le détail de leurs résultats, mais pour rendre plus concrètes les questions de méthodes et les limites des données.

Les questionnaires (A.)

Si on lit les manuels généraux (Tableau 2), la source principale de données sur les réseaux sociaux est le questionnaire, que l'on travaille sur de petits groupes ou de grands échantillons, des réseaux égocentrés ou « complets ». Il a fait l'objet de plus de publications méthodologiques que d'autres modes de collecte ; nous rassemblons et résumons ici leurs prescriptions, en nous appuyant sur une synthèse originale des débats ayant eu cours aux États-Unis sur les générateurs de noms utilisés dans le *General Social Survey*.

Réseaux égocentrés et « générateurs de noms »

Les questionnaires concernant des réseaux égocentrés visent à reconstituer l'entourage d'*ego*, les *alters* qui lui sont liés, afin de comparer la taille ou la composition des entourages selon différentes variables (L3). Pour cela, la méthode la plus courante est celle des « générateurs de noms ». Les questions que l'on appelle ainsi explicitent les types de relation que l'on cherche à étudier. L'enjeu de leur formulation est dès lors très important ; il faut, dans l'idéal, qu'elles soient comprises par tout le monde, et de la même manière. Les réponses à ces « générateurs » sont des listes de noms (ou prénoms, initiales, etc.) d'*alters* avec lesquels *ego* est en relation. Ces générateurs sont en général suivis d'autres questions appelées « interpréteurs de noms », qui demandent à *ego* des précisions sur les attributs des *alters* cités, sur les relations *ego-alter* (type de relation, ancienneté, etc.), voire sur les relations entre *alters* (sur les difficultés liées à ce dernier type de question, voir par exemple Marineau et Labianca, 2021).

Dans les grandes enquêtes par sondage, l'objectif est souvent d'approcher la définition d'un entourage immédiat et intime. Ainsi, les éditions de 1985 et 2004 de l'*American General Social Survey* (GSS) utilisaient un unique générateur de noms : « *De temps en temps, la plupart des gens discutent de problèmes importants avec d'autres personnes. Au cours des 6 derniers mois, qui sont les personnes avec qui vous avez discuté de problèmes importants pour vous ? Donnez-moi uniquement leur prénom ou leurs initiales.* » La comparaison des résultats aux deux dates semblait indiquer une décroissance du nombre moyen de confident·es et un accroissement de la part des personnes sans confident·e hors de leur famille (McPherson et al., 2006). Repris dans la presse, ce résultat a été remis en cause par d'autres sociologues qui ont pointé, à cette occasion, certaines difficultés et limites plus générales des générateurs de noms (Fischer, 2009 ; Paik et Sanchagrin, 2013). Cette controverse nous permet de les résumer en cinq précautions à prendre.

1. L'« effet enquêteur » est particulièrement fort (L2) : le nombre de noms cités est très variable selon la personne qui pose des questions, et notamment selon qu'il y a eu une relance ou non (« y a-t-il quelqu'un d'autre ? ») (Marsden, 2003). La baisse du nombre de personnes citées aux États-Unis est ainsi liée à une évolution de la formation et de l'attitude des enquêteurs et enquêtrices (Paik et Sanchagrin, 2013). L'auto-administration des

questionnaires en ligne nuit autrement à la qualité des données collectées (elle réduit notamment le nombre de noms cités) mais permet d'éviter cet « effet enquêteur » en homogénéisant les conditions de passation du questionnaire (Fischer et Bayham, 2019). Une autre solution peut être de standardiser les relances (Marin, 2004), en demandant un nombre prédéfini de fois « y a-t-il quelqu'un d'autre ? ». Il faut en tout cas éviter de contraindre la réponse par un nombre prédéfini de personnes à citer (Holland et Leinhardt, 1973). Cela peut faciliter l'enquête sur le moment, mais l'inégale répartition des relations est une réalité sociale : l'ignorer volontairement, forcer certain·es répondant·es à penser à plus de contacts ou à limiter leur liste, revient à déformer la structure des réseaux.

2. Les générateurs de noms sont des questions exigeantes, qui impliquent une fatigue pour la personne interrogée ; mieux vaut dès lors les placer le plus tôt possible dans le questionnaire, ce qui n'était pas le cas en 2004 aux États-Unis (Fischer, 2009) (L4). Il peut également être intéressant de réfléchir à des modes de passation en ligne qui présentent un aspect intuitif, voire ludique, en amenant les participant·es à dessiner leurs propres relations (Birkett et alii, 2021).
3. On l'a déjà dit, la formulation de la question est importante (L6). À l'origine, le générateur de noms utilisé provenait de l'enquête NCCS (*Northern California Community Study*) (voir Fischer, 1982), qui demandait de citer les personnes avec qui l'on parle de ses « problèmes personnels ». La formulation a été changée en « problèmes importants » pour augmenter le nombre de personnes citant plusieurs *alters* (Burt et alii, 1985). Mais le mot « important » peut être compris très différemment selon les personnes, notamment comme s'appliquant plutôt à des problèmes personnels ou non personnels. Dans un article au nom évocateur : « Cloning headless frogs and other important matters », Bearman et Parigi (2004) ont répliqué la question du GSS en demandant de préciser les « problèmes importants » auxquels pensaient les enquêtés. Certains pensent à des sujets généraux tels que l'éducation des enfants, ou des problèmes politiques, tandis que, d'autres évoquent des problèmes personnels liés au travail ou à la santé. Mais surtout, la moitié des répondants qui n'ont cité aucun *alter* considéraient simplement ne pas avoir de « problème important » à discuter : les sujets vraiment importants à leurs yeux – comme le contrôle des armes ou le clonage de grenouilles sans têtes – étant en dehors de leur domaine de compétence. De ce fait, il vaut mieux utiliser des termes précis liés à l'accès à des ressources (par exemple « une aide matérielle en cas d'urgence ») ou à des contextes relationnels bien spécifiés (par exemple « les personnes avec qui vous allez au restaurant ou avec qui vous partagez un dîner »). Si l'objectif est de délimiter l'ensemble de l'entourage, il faut utiliser *plusieurs* générateurs de noms : il paraît illusoire d'approcher cette notion avec une seule question (McCallister et Fischer, 1978).
4. Il faut bien garder en tête que la réponse à un questionnaire est un discours sur des relations et non une observation directe de leur fonctionnement (L2, L6). Small (2017) montre ainsi que les personnes que l'on cite dans un générateur de noms sur les confidences ne sont que très rarement celles vers qui on se tourne en pratique. Très souvent, on se confie à de simples connaissances et parfois à de parfait·es inconnu·es, car ce qui importe est avant tout le contexte de la relation – par exemple, un·e étudiant·e peut se tourner vers des étudiant·es du même âge, évoluant dans le même milieu, donc à même de comprendre sa situation. On peut aussi éviter ses proches pour ne pas les heurter ou pour conserver une certaine confidentialité. Un générateur de noms convient donc si on veut savoir qui la personne considère comme ses proches, mais guère pour mesurer des échanges sociaux effectifs. On peut toutefois approcher ce second cas en utilisant des questions factuelles, précises et bornées temporellement, par exemple : « à qui avez-vous parlé de vos problèmes financiers au cours des deux dernières semaines ? »

5. On ne pense souvent qu'à des générateurs de noms « positifs », évoquant des personnes qui procurent un soutien ou des ressources (L5). Mais, selon les questions de recherche, il peut être intéressant d'envisager aussi les personnes avec qui les relations sont difficiles ou qui « demandent beaucoup » à *ego* – une version des liens négatifs (Offer et Fischer, 2018). C'est un point à ne pas oublier si on ambitionne de parler de l'ensemble d'un entourage.

Un exemple de recherche tentant de tenir compte de ces impératifs est l'enquête RESTIC du Labex SMS réalisée en 2007 sur un échantillon de 709 personnes dans la région toulousaine (RESTIC, 2017). Le questionnaire est contraint par une double volonté de comparaison, avec une précédente enquête menée à Toulouse en 2001 et avec la série d'enquêtes NCCS évoquée plus haut (Fischer, 1982). Aux questions liminaires sur le profil sociodémographique et la trajectoire géographique d'*ego* ont été ajoutées des questions sur son usage d'Internet et des médias sociaux (L7). Suivent dix générateurs de noms (sans limite de nombre de noms cités), visant à couvrir des relations les plus diverses possible. Certains sont directement traduits de la NCCS : ils évoquent les personnes avec qui *ego* parle de ses problèmes professionnels, avec qui il ou elle partage des moments de sociabilité, mais aussi, par exemple, qui surveillent la maison en cas d'absence. D'autres sont spécifiques : par exemple sur les personnes avec qui *ego* parle de ses opinions politiques, ou les collègues qu'il ou elle voit en dehors du travail. Les « interpréteurs de noms » portent ensuite sur le rôle de la relation (famille, collègue, ami·e, voisin·e, conjoint·e, etc.) et son intensité (L6). Puis – concession à la faisabilité – c'est seulement sur dix noms cités par *ego* (les premiers nouveaux noms cités dans chacun des générateurs) que des questions plus précises sont posées : âge, niveau d'étude, profession, lieu de vie, ancienneté de la relation, contexte de la rencontre, etc. (L7). C'est également seulement pour ces dix noms qu'il est demandé à *ego* d'indiquer si *alter1* « connaît bien » *alter2*, *alter3*, et ainsi de suite, avec pour précision : « ils/elles sont capables de se contacter sans passer par vous ». L'enquête RESTIC propose ainsi une manière parmi bien d'autres d'équilibrer les précautions listées ci-dessus et les contraintes de temps et d'argent.

D'autres enquêtes étudient l'entourage social d'une personne en recourant à des questions plus simples, souvent du fait de contraintes de temps, mais aussi pour répondre à des problématiques précises. L'*International Social Survey Programme* (ISSP), une enquête auto-administrée réalisée dans un grand nombre de pays, qui comportait un volet sur les réseaux sociaux en 1989, 2001 et 2017, réalise un arbitrage différent en éliminant complètement la dimension nominative, pour proposer des questions en termes de nombres (par exemple : « En pensant à des gens sur votre lieu de travail, combien sont des ami(e)s proches ? ») ou de types de personnes. De la même manière, en sacrifiant la précision au profit de la faisabilité, le GSS de 1985 demandait si « en général, tous vos amis se connaissent entre eux » plutôt que d'interroger *ego* sur chaque lien entre *alters*. Lorsqu'on utilise ce type de questions pour savoir si *ego* peut avoir accès à certaines ressources via des *alters*, sans se préoccuper du nom ni des attributs des *alters*, on parle d'un « générateur de ressources » plutôt que de noms. On peut ainsi demander : « connaissez-vous quelqu'un qui peut réparer un vélo ? » ou « qui s'y connaît bien en matière de normes officielles », par exemple (Van der Gaag et Snijders, 2005).

Tout dépend donc des moyens disponibles, mais aussi des questions de recherche exactes. Autre exemple : plusieurs enquêtes ont tenté d'objectiver les relations entre professions sans passer par un générateur de nom, puis l'indication des attributs des *alters*, mais en demandant directement, par exemple dans l'ISSP 2017 : « connaissez-vous une personne qui exerce cette profession et dont vous êtes proche (vous la connaissez par son nom et assez bien pour la contacter) », question suivie d'une liste de professions (mécanicien automobile, avocat, etc.). Cet outil, introduit par Nan Lin, qui le qualifie de « générateur de positions », a pour objectif d'évaluer les inégalités d'accès à des ressources sociales dans différentes sociétés (Lin et alii, 2001). On sacrifie donc ici la connaissance générale de l'entourage d'*ego* (sa structure,

le type de relations qui y domine) et l'étude des ressources précises auxquelles il donne accès, au profit d'une entrée par la diversité sociale et le statut social des relations. Ce type de question implique de prédéfinir les différentes professions, ce qui ne permet donc pas d'avoir une vision exhaustive de l'entourage du répondant, et pose également des difficultés dans les comparaisons dans le temps et dans l'espace (Marsden, 2011). Comme il n'est jamais possible de tout savoir – et de manière parfaitement comparable entre cas – des entourages des personnes ou de leur capital social, il est important d'adapter la formulation des questions à la problématique précise de l'enquête (Hällsten et alii, 2015).

Réseaux « complets » et questionnaires sociométriques

Qu'elles utilisent des générateurs de noms, des générateurs de ressources ou de positions, les enquêtes sur des réseaux égocentrés décrivent donc avant tout les liens entre *egos* et *alters*. L'ajout de questions sur les liens entre *alters*, permettant de décrire la structure du réseau, est loin d'être systématique, car il est particulièrement coûteux en temps. Au contraire, la structure d'ensemble du réseau est le principal élément que cherchent à décrire les enquêtes sur des réseaux « complets », celles dans lesquelles on part d'une liste d'individus ou d'organisations pour voir, lien potentiel par lien potentiel, s'il existe et comment le qualifier. La technique spécifique d'enquête est donc le questionnaire sociométrique, introduit par le psychologue Jacob Moreno (1953) et son équipe. Il s'agit de questionner chaque individu sur ses liens avec tous les autres, un par un. Emmanuel Lazega (2014) en présente les étapes et les difficultés à partir d'un exemple. Nous les reprenons ici brièvement en suivant une autre enquête, qui a la particularité d'envisager à la fois des relations entre individus et entre organisations (Favre et Brailly, 2015). Elle avait pour but d'étudier la construction sociale d'un marché africain des programmes de télévision à partir d'une analyse de réseaux des participant·es de salons ayant eu lieu en 2010 à Dakar, 2011 à Accra et 2012 à Johannesburg. Dans ces salons se rencontrent acheteurs (majoritairement des chaînes de télévision africaines) et vendeurs (majoritairement des distributeurs internationaux) de programmes de télévision. La réalisation de l'enquête était contrainte par la durée relativement brève de ces événements.

La première étape, en amont de l'élaboration du questionnaire proprement dit, est la définition de la frontière du groupe à étudier (L3). Elle influence le résultat obtenu, bien plus que dans le cas de réseaux égocentrés. Le cas des salons pouvait sembler simple, avec l'inclusion de toutes les personnes présentes – vendeurs et acheteurs de programmes – et seulement elles. Mais en réalité, d'autres personnes sont présentes : organisateurs du salon, services de coopération internationale, producteurs locaux à la recherche de financements, etc. En outre, bien des présents effectifs ne sont pas sur la liste officielle d'inscription. Il fallait donc revenir à la question de recherche pour trancher : ces acteurs jouent-ils un rôle dans la construction d'un marché africain des programmes de télévision ? En pratique, la réponse – inclure tou·tes les participant·es, pas seulement les vendeur·ses et acheteur·ses –, appuyée sur une enquête préliminaire, a toutefois fait évoluer la question, vers la construction d'un milieu social et non pas uniquement d'un marché.

Cette enquête préliminaire, reposant sur 62 entretiens semi-directifs auprès de participant·es, a été encore plus déterminante pour les deux étapes suivantes : la définition des relations à étudier et la formulation précise des questions à leur sujet (L2). Le but des entretiens était de découvrir le milieu de la distribution de programmes en Afrique, ses enjeux, de décomposer les tâches des commerciaux, d'identifier leurs interdépendances et, tout particulièrement, les ressources dont la circulation est nécessaire au milieu : pourquoi les acteurs ont-ils besoin les uns des autres, doivent-ils faire appel à leurs collègues, leurs clients ou leurs concurrents pour faire leur métier ? Les entretiens ont notamment permis de faire

émerger l'importance des recommandations et autres moyens de s'informer de la fiabilité des partenaires.

Comment, à partir de là, construire des questions précises ? Comme pour les générateurs de noms, il faut d'abord arbitrer entre la multiplication des questions, qui permet notamment de comparer la structure de différents types de liens, et la nécessité de limiter le temps de passation. Dans le cas des salons, les contraintes liées à la disponibilité des enquêté·es ne permettaient pas de poser plus de deux questions sociométriques. La première interrogeait les répondant·es sur les personnes avec qui ils ou elles avaient discuté lors du salon et dont ils ou elles avaient obtenu une information utile pour leur travail. Cette question énumérait cinq exemples d'informations, ce qui permettait d'espérer qu'elle soit comprise par tou·tes à peu près de la même manière : des informations sur un·e concurrent·e, sur les tendances du marché, sur les programmes à la mode, sur la fiabilité des client·es et des informations procurant des opportunités commerciales. La seconde question, portant sur les liens entre organisations, les interrogeait sur les sociétés avec lesquelles la leur avait conclu un contrat lors du salon ou lors de l'année précédente (on retrouve, comme pour les générateurs de noms, la nécessité de préciser une période à prendre en compte). Il faut noter que si les questionnaires sociométriques sont, le plus souvent, remplis de manière binaire (la relation ou interaction visée est présente ou absente), rien n'interdit en principe de proposer de la qualifier (comme fréquente ou non, de valence positive ou négative, etc.) (L6).

Pour l'enquête sur les salons, comme c'est souvent le cas pour les questionnaires sociométriques, la majorité des questionnaires ont été remplis par les répondant·es en face à face avec un·e enquêteur·trice (les autres, par défaut, par téléphone sur support électronique). La liste des participant·es (personnes ou sociétés) était soumise aux répondants, qui devaient cocher les cas où il y avait eu une discussion ou un contrat. Il est important de mettre en forme cette liste pour faciliter la passation du questionnaire. Ainsi, dans le cas des salons, elle comprenait près de 300 personnes : une ligne sur deux a été grisée et les personnes de la même entreprise apparaissaient dans une même case. Il est important de prendre le temps de réaliser des tests pour adapter cette présentation.

Le principal message à retenir concernant l'étude de réseaux « complets » est sans doute le fait qu'une enquête préalable (par entretiens semi-directifs, observation et/ou sur documents) s'impose pour déterminer des frontières pertinentes du groupe, une population à étudier et une formulation des questions adaptée à la fois à la problématique et aux personnes étudiées. Ces autres méthodes d'enquête, toutefois, ne sont pas seulement un préalable à l'élaboration de questionnaires. Dans d'autres études, elles constituent des sources plus directes de données – y compris de données systématiques et quantifiables – sur les réseaux sociaux.

Les entretiens semi-directifs (B.)

Cette méthode est bien plus rarement discutée dans les manuels d'ARS. Elle a pourtant été employée par Mark Granovetter (1974) dans son étude sur l'obtention d'un emploi et reprise depuis sur des thèmes divers, comme les entrepreneurs (Coviello, 2005) ou les réseaux personnels (Bidart et alii, 2011). L'objectif est d'obtenir d'*ego*, au cours d'un entretien souvent long, un récit qui l'associe à des *alters*, à une période précise et à une ou plusieurs ressources échangées – il s'agit souvent d'un récit de « chaînes relationnelles » (Killworth et alii, 2006). L'enjeu est de collecter un matériau narratif pour le transformer ensuite en variables décrivant des entités et des liens, dans l'optique de typifier les réseaux à l'aide des indicateurs quantitatifs classiques, éventuellement accompagnés d'une visualisation.

L'usage relativement limité de ce type d'entretiens tient sans doute à leur coût en termes de temps : il ne permet pas de collecter des données sur des grands échantillons et nécessite un lourd travail de mise en forme *a posteriori* du « récit ». Ce mode de recueil a été systématisé en

France sous l'appellation de « narrations quantifiées » par Michel Grossetti (2011) et appliquée à des thèmes variés : les relations science-industrie (Grossetti et Bès, 2001), la création d'entreprises (Chauvac, 2016), le recours aux médecins par les malades du cancer (Defossez, 2014), les inventeurs (Cloutier, 2014), etc. Cette approche n'est centrée ni sur des réseaux personnels, ni sur des réseaux complets, mais sur des séquences temporelles et relationnelles d'accès à des ressources (L3). On ne demande pas aux *egos* de restituer en priorité des listes de noms (et de commenter leurs relations), mais de reconstituer la manière dont ils ou elles ont eu accès à telle ou telle ressource, en passant ou non par des personnes.

C'est une méthode fondée sur le recueil de récits détaillés et croisés des acteurs à propos de la manière dont ils et elles s'y sont pris pour trouver une ressource précise (contrat, brevet, argent, client, information, etc.). Les entretiens visent à produire un récit chronologique des différentes démarches relationnelles effectuées par les acteurs, une narration séquencée du processus social ayant permis l'obtention de ressources indispensables à l'activité étudiée (invention, recherche d'emploi, création d'entreprise, etc.). Cela est permis par des questions comme : « Vous me dites que votre premier problème, avant de créer votre entreprise, était de trouver un comptable afin de réaliser une petite étude de marché, comment avez-vous fait pour le chercher ? ». L'enquêté·e donne alors plusieurs réponses que le chercheur ou la chercheuse va d'une part situer dans le temps et d'autre part catégoriser en fonction d'idéaux-types qui permettent de styliser, comparer et même quantifier des récits. Par exemple, on peut distinguer les séquences qui s'appuient sur des dispositifs institutionnels (comme des aides publiques), sur des fonctionnements marchands (répondre à une annonce) ou sur des relations entre personnes (obtenir un contact par un voisin). On peut également compter, pour chaque séquence, le nombre de personnes mobilisées. Ainsi, alors que d'autres méthodes visent à décrire la forme des relations entre personnes, ici, il s'agit plutôt de distinguer la part des relations interpersonnelles, par rapport à d'autres manières d'accéder à des ressources, dans le contexte de différents processus.

Le grand intérêt des entretiens semi-directifs en général, par rapport aux simples questionnaires, est bien sûr de permettre de qualifier plus précisément, dans les catégories de la personne enquêtée, aussi bien les liens que les entités (L6, L7). La possibilité de rebondir sur les réponses paraît également plutôt favorable pour le recueil d'informations sur les liens considérés comme négatifs par les personnes, ainsi que sur la temporalité des liens (Chauvac et alii, 2014) (L5). La limite, symétrique, est qu'ils prennent plus de temps et donc ne permettent que d'étudier des échantillons ou groupes plus petits et des processus sociaux circonscrits dans le temps. Quant aux entretiens visant à obtenir des narrations sur des chaînes relationnelles, par définition, ils ne permettent pas d'obtenir de récits sur les liens absents (L4) : l'objectif n'est pas ici de reconstituer un réseau « complet » (sociocentré) ou un réseau égocentré, mais d'obtenir une série de liens articulés entre eux (on peut parler d'approche lien-centrée). La non-prise en compte des échecs (même si certains des travaux précités sont attentifs aux liens non utilisés, voire négatifs) est une autre limite de la méthode des narrations quantifiées.

L'expérimentation (C.)

Nous présentons un peu plus longuement la construction de données par expérimentation, car elle est plus rarement évoquée dans les textes généraux sur l'ARS, comme sur les méthodes en sociologie en général. Elle peut pourtant être employée à des échelles variées. Nous revenons d'abord sur une étude pionnière souvent citée pour ses résultats généraux, mais dont le *design* précis et la postérité méthodologique n'ont, à notre connaissance, pas été discutés depuis. Puis nous présentons le retour d'expérience de l'une de nous.

Des données prouvant l'existence de grands mondes

Dans les années 1960, le psychologue social Stanley Milgram lance une série d'expériences pour tester l'assertion selon laquelle « le monde est décidément petit », traduite en probabilité forte que deux personnes prises au hasard dans le monde se connaissent mutuellement. Reformulant sa question, le chercheur (1967, 62) écrit : « *given any two people in the world, person X and person Z, how many intermediate acquaintance links are needed before X and Z are connected?* ». Il lance une première expérience peu concluante entre un échantillon de personnes situées dans la ville de Wichita au Kansas et une étudiante d'une école religieuse, située à Cambridge (Massachusetts), comme cible à atteindre. En effet, il cherche d'abord à voir si la distance géographique est un obstacle à la mise en contact entre deux personnes qui ne se connaissent pas. De cette étude, il tire quelques résultats préliminaires : la forte homophilie de sexe dans les envois et le recours majoritaire aux relations d'amitié et de connaissance par rapport aux liens familiaux.

Peu de mois après, il récidive avec la collaboration de Jeffrey Travers en choisissant 296 autres volontaires⁴ appartenant à trois sous-populations distinctes (le tiers habitent dans le Nebraska, le tiers habitent dans le Nebraska et sont porteurs d'actions et le tiers habitent à Boston) et une autre personne cible : un agent de change de Boston. Il s'agit de comparer les chaînes reposant sur la proximité géographique et celles reposant sur une proximité d'activité, ou d'intérêt pour la finance (Travers et Milgram, 1969). Le dispositif de recherche est astucieux : chaque volontaire dispose d'une chemise contenant quatre éléments : premièrement, une description de l'étude, un accord de participation et l'exposé des règles de participation (faire passer la lettre à une personne dont on pense qu'elle pourrait connaître la cible), deuxièmement, le nom de la cible et des informations la concernant (l'article publié ne dit malheureusement pas lesquelles), troisièmement, un tableau de suivi dans lequel chaque participant·e appose son nom et quatrièmement, une série de cartes à remplir par chaque participant·e et à renvoyer aux expérimentateurs. Ce protocole permet donc aux deux chercheurs de suivre le cheminement de la « lettre » et quelques caractéristiques des personnes qui la font circuler (sexe, âge, profession, localisation) (L7). Les résultats sont assez pauvres : seuls 217 des premier·es volontaires sur les 296 recruté·es renvoient la lettre, seules 64 lettres atteignent leurs cibles, soit 21 % ou 29 % selon la manière de calculer le ratio de « succès ». Dans la mesure où près de 80 % des lettres ne sont jamais arrivées, comme l'écrit avec humour Judith Kleinfeld (2002) : « *it could be a big world after all* ». Il n'empêche que la littérature ultérieure retient que le monde est petit et qu'il faut moins de six intermédiaires pour relier deux inconnus.

Après cette première expérience sociale marquante, d'autres sociologues ont lancé, à partir des années 1970, des applications de la même méthode, dite du « petit monde ». Elles se scindent en deux groupes : celles centrées sur les questions de stratification sociale (Korte et Milgram, 1970 ; Lin et alii, 1978) et celles centrées sur le fonctionnement de certaines organisations, notamment les universités. Si les premières ont affiché des taux de complétude proches de l'expérience mère (entre 13 et 33 %), les auteurs en ont quand même conclu que les « Noirs étaient des cibles plus difficiles à atteindre que les Blancs » ou bien que les personnes d'un milieu social aisé réussissaient bien plus facilement à « descendre la pyramide sociale » que celles d'un milieu populaire à la remonter. Dans le second groupe, on trouve par exemple l'étude menée au sein d'une Université de management de Boston afin de comparer la capacité des étudiant·es à atteindre un service administratif en particulier (Stevenson et alii, 1997). 60

⁴ L'article ne précise pas si les volontaires sont rémunéré·es mais, comme c'est souvent le cas dans les enquêtes états-uniennes, il y a fort à parier que c'est aussi le cas. Dans ce type d'expérimentation, c'est un peu gênant car cela peut performer le résultat, à savoir l'obtention d'une chaîne complète, qui constitue la donnée à analyser.

d'entre eux et elles, aux différents stades de la formation, ont été sollicités pour contacter, de proche en proche, une cible choisie parmi les salarié·es exerçant des fonctions administratives. Comme seuls 16 étudiant·es ont réussi, toutes en deuxième ou troisième cycle, l'étude a conclu à la difficulté pour les étudiant·es de première cycle – en particulier les garçons – d'obtenir des informations administratives.

Au début des années 2000, alors qu'Internet offre la possibilité d'envoyer très rapidement et très abondamment des messages, des chercheurs états-uniens (Dodds et alii, 2003) reproduisent l'expérience de Milgram sur une échelle bien plus grande. 98 847 personnes se portent volontaires pour « faire passer un message jusqu'à une personne cible en utilisant une connaissance qu'elles considèrent comme plus proche de la cible qu'elles-mêmes ». Les chercheurs ont sélectionné 18 personnes cibles aux caractéristiques différentes, ont réparti leur effectif de volontaires entre elles et ont donné aux « starters » les informations suivantes sur la personne cible : le nom, la ville de résidence, le métier, le diplôme. Le taux de chaînes complètes, c'est-à-dire celles qui atteignent la cible, est extrêmement bas (en moyenne de 0,4 % avec un écart-type de 0,37 %) : si aucun des 8 151 volontaires ne parvient jusqu'au chômeur indonésien, le professeur états-unien habitant à Ithaca, la plus « populaire » des cibles, n'est atteint que par 168 chaînes parmi celles lancées par les 5 855 premiers individus qui l'avaient pour cible, soit un taux de complétude de 2,87 %. Malgré cela, les auteurs de l'étude tirent quelques nouveaux enseignements : premièrement, leurs chaînes ne passent pas par des personnes clés (*hubs*) incontournables, alors qu'elles étaient sollicitées fortement dans l'expérience de Travers et Milgram, et deuxièmement, les relations professionnelles, dites « faibles », sont plus souvent activées que les relations fortes (familiales et amicales).

Dans le même numéro de la revue *Science*, Mark Granovetter (2003) commente cette étude de manière assez critique en soulevant trois problèmes complémentaires. D'abord, l'expérimentateur, en donnant les caractéristiques de la cible (nom, lieu de vie, emploi), influence la manière dont les individus sollicitent leurs réseaux : en effet, dans leur vie de tous les jours, ils s'appuient sur des informations plus subtiles (L2). Ensuite, les répondants devaient indiquer un seul et unique motif pour lequel ils avaient choisi un destinataire. Cela les a amenés à privilégier le lieu de vie ou l'emploi du destinataire au détriment d'autres dimensions, comme la taille de son carnet d'adresse par exemple. Cela peut expliquer que des personnes sur-sollicitées (des *hubs*) repérées dans d'autres études ne sont pas ressorties dans cette expérience. Enfin, il est nécessaire de s'intéresser à ce qui circule dans la prise de contact (information, virus, maladie, etc.) afin de ne pas perdre de vue l'enjeu de la fabrication d'une chaîne et de ne pas surinterpréter les résultats (L1, L6) : pourquoi les participant·es à ces expérimentations ont-ils et elles fait circuler le message ? pour recevoir la récompense financière promise par les chercheurs ? pour gagner en reconnaissance auprès des *alters* ? pour rendre service ? etc.

Cela revient à souligner que ces enquêtes visant à une généralisation extrême (dire quelque chose du monde social dans son entier) n'ont pas prêté une assez grande attention à leurs propres biais, pas assez clairement défini ce qu'elles observaient ou non, notamment par manque d'intérêt pour la qualification précise des liens en cause. Les enquêtes par expérimentation portant sur une problématique précise dans un milieu social ou une organisation spécifiques échappent plus facilement à ces travers (voir par exemple Singh et alii, 2010 au sein d'une entreprise).

Obtenir des données partielles en suivant le cheminement d'un e-mail

Comment peut-on concrètement procéder ? Prenons l'exemple d'une enquête menée en 2007 par Marie-Pierre Bès (2011). Il s'agit de tester l'existence de réseaux solidaires entre diplômé·es d'une école d'ingénieur, dans laquelle l'enquêtrice est enseignante. Elle a enrôlé un étudiant de cette école, qui a contacté les ancien·es élèves, s'est chargé de lancer l'expérimentation et de

collecter les données. La problématique est de savoir si les diplômé·es d'une école d'ingénieurs gardent des contacts, longtemps après être sorti·es de leur école et si oui, pourquoi. Est-ce la trace d'une sociabilité amicale, apparue pendant leurs années d'étude, ou bien celle d'une forte concentration d'ingénieur·es au profil identique dans les mêmes entreprises ? Ou bien la conséquence de l'existence d'un fort sentiment d'appartenance à la communauté des diplômé·es, créant un attachement entre deux diplômé·es qui ne se connaissent pas directement ?

La méthode utilisée s'appuie sur les expérimentations évoquées ci-dessus. Il s'agit là aussi de faire circuler un message pour voir par quelles personnes il passe et ainsi comprendre la mobilisation de cercles sociaux. Mais l'expérience lancée ne comporte pas de cible à atteindre, puisque la consigne envoyée par l'étudiant aux diplômé·es sélectionné·es dans l'annuaire est : « en utilisant votre carnet d'adresses, transmettez ce mail à des diplômés de votre école et mettez-moi en copie, cela m'aidera ». Les participant·es à cette expérience ont donc, chacun·e, transmis à d'autres individus, dont ils et elles pensaient qu'ils ou elles étaient diplômé·e de la même école, l'e-mail initial envoyé par l'étudiant, mentionnant que leur participation l'aiderait à finaliser son projet d'étude de deuxième année, qu'ils ont dû eux-mêmes conduire quelques années auparavant.

Les données de cette recherche ne se limitent pas à celles produites par l'expérimentation au sens strict. Il est possible de caractériser les participant·es aux chaînes par des attributs à la fois à partir des données expérimentales (une partie de leur adresse e-mail indique leur employeur) et d'une source écrite utilisée pour lancer l'expérience : l'annuaire des diplômé·es (L7). Sans cette double collecte, les listes de noms engendrées ne seraient pas très utiles. En effet, ce sont les attributs (employeur, genre, année de diplomation) qui permettent de comprendre ce qui porte ou non la conservation d'une sociabilité après la sortie de l'école d'ingénieurs.

Sur les 61 chaînes lancées, seules 10 ont démarré⁵, dont 7 courtes et étroites, cantonnées à quelques ami·es et conjoint·es, 2 courtes mais larges, qui se sont diffusées dans un petit milieu professionnel, et une très longue (21 séquences) et large (15 destinataires en moyenne), qui a alterné les relations personnelles et anciennes de camaraderie et des nouvelles relations professionnelles.

Dans cette recherche, l'envoi et la réception de l'e-mail sont interprétés comme l'indice de l'existence d'un lien social. L'enquêtrice ne dispose d'aucune information directe sur la nature de cette relation (simple collègue ou amie, conjoint ou vague connaissance, voisin ou sous-traitant quelconque), mais utilise les attributs des personnes pour faire des hypothèses à ce sujet (L6). Ces hypothèses dépendent de sa problématique de départ, visant à distinguer un « effet professionnel » d'un « effet campus ». L'attribut lié à l'employeur a permis de qualifier ou non la relation entre l'émetteur et le récepteur comme un lien professionnel. Mais il est important de se souvenir, pour l'interpréter, que cette variable n'est pas un indicateur de la nature de la collaboration professionnelle elle-même, donc de la relation (L1). Elle signifie seulement qu'une personne d'une entreprise a renvoyé le message à une autre travaillant dans la même, car elle savait que cette dernière était diplômée de la même école que lui. Cette diffusion a probablement été facilitée par le fait de disposer d'un annuaire électronique partagé, générant facilement une adresse à partir d'un prénom et d'un nom. L'année de diplomation, indiquée dans l'annuaire, permet, elle, de créer un lien de co-appartenance entre deux individus de la même promotion (chacun compte entre 50 et 100 élèves). On peut faire l'hypothèse que ce lien indique une certaine camaraderie, mais l'auteure n'a pas enquêté sur son contenu vécu dans chaque cas. Sa position d'observatrice participante comme professeure dans l'école lui

⁵ On ne sait jamais si cela est dû à un manque de volonté et de disponibilité des participant·es ou à des questions plus matérielles : avoir changé d'adresse e-mail, ne plus consulter celle indiquée dans l'annuaire, etc.

permet en revanche d'étendre la même hypothèse d'une possible interaction directe entre les élèves à ceux et celles situés dans les deux promotions antérieure et ultérieure, compte tenu des activités de sociabilité existant sur ce campus. Ainsi, c'est la bonne connaissance antérieure de ce que peut être la substance des relations sur le terrain qui permet d'enrichir les données issues de l'expérimentation et de les interpréter sans les surinterpréter.

L'observation (D.)

Que l'on procède par questionnaire, expérimentation ou entretien, on place d'une manière ou d'une autre les enquêtés dans une situation artificielle, hors de leur quotidien, où on leur demande de dire quelque chose de leurs interactions et relations (ou, dans le cas de l'expérimentation, d'agir sur elles d'une manière choisie par l'enquêteur ou l'enquêtrice). Cette manière de faire a des avantages et des inconvénients : il faut bien rester conscient·e qu'on travaille alors sur des discours sur des liens, plus ou moins influencés par le mode d'enquête, et pas sur « des liens » dans l'absolu (L1).

En contrepartie, il peut être intéressant de recourir à l'observation, pour enregistrer ce qui se passe devant l'enquêteur ou l'enquêtrice, dans certaines circonstances, plutôt que ce qui s'est passé, selon les dires des enquêtés, lors d'une interaction antérieure. Cela n'élimine évidemment pas l'effet enquêteur·trice, mais en change les modalités (L2). L'utopie serait d'observer chaque interaction, sans interférer avec elles. En réalité, on sait bien que l'observation influence souvent ce qui est observé, et surtout qu'on n'observe pas tout également : on fait toujours des choix.

L'observation est régulièrement évoquée au détour de livres ou d'articles pratiquant l'ARS, mais en général en combinaison avec d'autres méthodes, et surtout comme approche exploratoire plutôt que comme source de données systématiques ou quantifiables. Le travail fondateur, très souvent cité, de John Barnes (1954) sur une île norvégienne, outre qu'il ne repose pas exclusivement sur l'observation, propose ainsi une théorisation qui ne passe pas par l'analyse de données systématiques (sur le contexte de cette recherche, voir Mercklé, 2013). Plusieurs des études intensives de petits groupes qui ont été reprises par les promoteurs d'une ARS plus formalisée ont mobilisé entre autres l'observation, mais elle n'est jamais la source directe des matrices de données analysées (White et alii, 1976). La recherche, fameuse en sociologie économique, de Wayne Baker (1984) sur la structure sociale du commerce d'options dans une bourse fournit un exemple classique de méthodes multiples utilisées ainsi de manière séquentielle. L'observation participante est son principal moyen de découverte du terrain (complété par des entretiens informels). Cela lui permet de décider quel réseau de transactions extraire à partir des données écrites de la bourse : c'est ce réseau qu'il analyse quantitativement. Enfin, Baker conduit de nouveaux entretiens : il fait réagir les enquêtés aux structures de réseau qu'il a mises au jour.

À cette pratique fréquente de l'observation comme un préalable permettant de construire un questionnaire pour mesurer systématiquement les relations, on pourrait pourtant opposer le constat que les observations d'interactions et les discours à leur sujet peuvent se contredire (L2). En effet, dans le cadre des premières enquêtes de psychologie sociale sur des jeux entre enfants, Elizabeth Hagman remarque cette contradiction dès 1933 (Freeman, 2004 : 21). Depuis, le problème a été régulièrement redécouvert, avec notamment une série d'articles centrés sur les données concernant des communications, téléphoniques par exemple (ces articles sont également pionniers en ce qui concerne l'usage de traces numériques d'interaction : Bernard et alii, 1982). Mais dans l'ensemble, il paraît plutôt oublié, au profit d'un usage routinier du questionnaire. Par exemple, Ann Mische (2003 : 273) ne parle que de manière programmatique d'*observational coding*, c'est-à-dire d'une observatrice qui noterait systématiquement certains aspects de conversations entre activistes. Pourtant, ce type d'observation quantifiée était mené

dès les années 1930, dans des bureaux aussi bien que des cours d'école (Freeman, 2004 : 49). Dans la littérature récente en ARS, cette source de données est devenue très rare.

L'étude de la sociabilité au sein d'un club de *fitness* par Nick Crossley (2008) fait exception. En plus d'éléments formulés de manière narrative, il présente des analyses formalisées sur un « réseau complet » qui se fondent exclusivement sur des observations participantes. Plusieurs types de liens sont pris en compte : la participation commune à trois types d'activités, mais aussi deux types de relations inter-individuelles (L6). À chaque fois, la définition de la relation est d'autant plus précise qu'elle est fondée sur un travail ethnographique. L'une des relations, en outre, ne pourrait pas être décrite autrement : il s'agit du fait, pour deux membres du club, à la fois de se rencontrer (quelle que soit la raison) également hors du club et de le manifester par des « signes de liens » au sein du club. Crossley ne systématise pas, en revanche, l'étude des liens négatifs, mais il est clair que sa méthode permettrait de le faire (L5). Cet exemple montre la richesse de l'observation – participante en particulier – pour documenter la substance des relations. Elle permet aussi de connaître des attributs individuels qui seraient peut-être plus difficiles à documenter autrement, comme le degré d'engagement dans la vie familiale. Mais dans le même temps, Crossley n'a qu'une idée assez vague des revenus de ses pair-es – sans doute un sujet de discussion relativement tabou (L7). Enfin, bien évidemment, sa fenêtre d'observation est réduite ; il souligne ainsi qu'il aurait sans doute vu des choses très différentes s'il avait fréquenté le club en journée plutôt qu'en soirée (L3).

L'observation peut ainsi être une source de données pertinentes pour l'ARS. Ce que nous avons en vue ici, c'est une observation directe, accompagnée de prise immédiate (ou presque) de notes. Dès lors qu'un enregistrement intervient, en revanche – et en particulier s'il n'est pas guidé par les choix d'un enquêteur ou une enquêtrice présent-e –, c'est à un autre type de source de données qu'on a affaire : une trace numérique (vidéo, ou audio comme chez Corman et alii – 2021 –, ou issue d'un autre type de capteur). L'enregistrement automatique permet certes d'obtenir une plus grande quantité de données que l'observation, mais il présente le risque d'appauvrir, en contrepartie, les informations sur les interactions captées : on peut ainsi savoir que des enfants se font face, à moins d'1,5 mètre, dans leur classe, mais absolument pas ce qu'ils et elles échangent (Stehlé et alii, 2011) (L6). La méthode, qui se présente comme moins biaisée que celle des entretiens, présente donc plutôt des avantages et inconvénients différents – assez proches, en réalité, de ceux des sources écrites, qui sont d'autres traces de pratiques.

Les traces de pratiques : sources écrites et numériques (E. & F.)

Le recours à des documents constituant des traces de pratiques⁶ plus ou moins anciennes peut être considéré comme une variante des situations d'observation. La différence est que dans ce cas, l'effet enquêteur·trice n'intervient plus dans la production de la source, mais seulement dans la construction des données, lors des choix que nous avons évoqués dans la première partie de cet article : quel échantillon ou groupe exact étudier, quels renseignements sur les entités aller chercher dans la même source ou une autre, comment catégoriser les types de liens, etc. La responsabilité de cette construction des données demeure : deux chercheur·ses n'élaboreront pas les mêmes données (et moins encore les mêmes graphes, indicateurs, interprétations) à partir des mêmes sources, et c'est une richesse pour la recherche (la construction des données dépend de la problématique), pas un problème – tant que l'on s'en rend compte.

⁶ Nous utilisons ici le terme « pratique » dans le sens du langage courant, pour décrire toute forme d'activité susceptible de laisser des traces.

Cela permet notamment de revenir à une source pour reconstruire pour partie les données un peu différemment au cours d'une recherche, chose bien plus difficile à faire, voire impossible, dans le cas de questionnaires, entretiens, expérimentations ou observations. En revanche, notons que lorsqu'on réutilise les données produites par d'autres personnes, ou par soi-même, lors d'une enquête précédente employant ces méthodes, on se retrouve de fait dans la même situation que lorsqu'on utilise une archive, des données *scrapées* depuis le web ou produites par des capteurs : il faut faire une « critique de la source », comprendre ce qu'elle documente et ce qu'elle ne documente pas, qui l'a produite, dans quel objectif et avec quelles catégories, avant de l'utiliser (L1).

Toutes sortes d'interactions et de relations sont susceptibles de laisser des traces (écrites, numériques, ou encore dessinées, sculptées, etc.) qui deviennent ensuite utilisables comme sources pour la recherche. Parfois, ces traces découlent d'une obligation légale d'enregistrement (d'un mariage, d'une vente, etc.) ou du fait que certaines personnes réalisent un enregistrement pour assurer leurs droits (brevets, signatures d'œuvres d'art, etc.). Des organisations produisent des listes de membres ou de dirigeants, dans divers buts (légitimation, communication, etc.) ; on peut interpréter ces listes comme des liens entre individus et organisations. D'autres traces de pratiques fournissent, plutôt que des listes assez sèches de liens similaires entre eux, un discours précis sur chaque lien. C'est notamment le cas des sources narratives, ou encore des correspondances, qui détaillent des relations d'amitié, d'affinité intellectuelle, d'opposition politique, etc. (L2) (pour plus de détails sur la construction de données à partir de ces grands types de traces, voir Lemerrier, à paraître). On retrouve ces mêmes types de traces dans des documents nativement numériques : les principes de critique des sources correspondants s'appliquent également au *big data*. En effet, les traces laissées « automatiquement » par des interactions en ligne ne nous disent pas tout de ces interactions (L2). Tout d'abord, elles sont inégalement accessibles pour la recherche : Facebook a par exemple mis à disposition, un temps, des éléments que l'entreprise refuse maintenant de partager (pour un exemple d'utilisation de ce type de données, voir Fitzhugh et Butts, 2018). Cet accès dépend à la fois des choix des individus que nous souhaitons étudier (ce qu'ils et elles savent et veulent mettre en ligne plus ou moins publiquement) et de ceux des plates-formes. Il ne s'agit pas seulement du choix de leurs dirigeant·es d'ouvrir ou non leurs données à la recherche. Ces données ont également été pré-construites : leurs formats dépendent de choix effectués par les programmeur·ses. Ainsi, le *big data* ne documente que les éléments mis en ligne par les personnes que l'on veut étudier *et* que les programmeur·ses qui ont mis au point l'enregistrement « automatique » trouvaient importants ou utiles. Il faut donc essayer de comprendre ces choix de publication, mais aussi d'enregistrement et ce qui est laissé de côté. Comme dans le cas de données issues d'archives, on peut progresser dans cette compréhension soit par l'observation des données elles-mêmes (notamment lors de la phase cruciale dite de « nettoyage » pendant laquelle on les remet en forme), soit par le recours à des sources complémentaires. Une fois que l'on connaît ces biais, on peut essayer de les corriger en cherchant des informations complémentaires dans d'autres sources, ou au moins proposer une interprétation des données qui en tient compte.

Le recours à des traces d'interactions ou de relations, qu'elles soient nativement numériques ou non, comme sources pour étudier les réseaux sociaux présente plusieurs avantages. C'est d'abord la seule solution si l'on veut parler d'interactions ou de relations entre des personnes décédées, que l'on ne peut plus interroger. Si ces personnes sont encore vivantes mais que les liens concernés sont anciens, les documents peuvent parfois compléter ou suppléer une mémoire défaillante (mais les personnes peuvent aussi se souvenir de liens qui n'ont pas laissé de traces...) (L2). Plus généralement, le recours à des enregistrements de communications montre que les traces donnent, en la matière, des résultats plus complets que les questionnaires, notamment parce que lorsqu'il y a dans l'ensemble beaucoup de communications, bien des

interactions sont oubliées (L4) (Corman et alii, 2021). De la même manière, il est possible que des liens qui gênent les personnes concernées, et dont elles ne souhaiteraient pas parler dans le cadre d'un entretien ou d'un questionnaire, aient laissé des traces (L5). C'est notamment ce qu'espèrent les personnes qui utilisent des documents pour étudier des activités illégales (Bright et alii, 2021). Les études sur des mouvements d'opposition politique durement réprimés dans le passé, comme celle de Marten Düring (2013, 2015) sur le sauvetage de Juif·ves dans l'Allemagne nazie ou de Maryjane Osa (2003) sur les mouvements d'opposition en Pologne dans la période communiste, donnent un bon exemple d'un usage prudent d'une grande variété de sources, dont des entretiens (y compris la réutilisation d'entretiens menés par d'autres chercheur·ses), des sources du mouvement étudié et de la répression, pour comprendre les logiques relationnelles qui sous-tendaient ces mobilisations.

Deux autres avantages des traces de pratiques expliquent qu'elles aient été la source de bien des travaux pionniers de l'ARS, relevant de la sociologie historique, et qu'elles soient aujourd'hui largement reprises par les études des interactions numériques. D'une part, elles permettent une observation sur une certaine durée et les interactions ou relations y sont souvent datées – alors qu'il est difficile et coûteux de recourir à plusieurs vagues d'enquête par entretien ou questionnaire et qu'il n'est pas toujours évident d'obtenir une date pour une interaction ou relation mentionnée en entretien (Bidart et alii, 2011) (L5). Ainsi, les données issues de traces de pratiques peuvent se prêter à une étude du changement, sur un temps plus ou moins long.

D'autre part, avec certains types de traces, il est relativement peu coûteux d'obtenir des informations sur un grand nombre d'entités – du moins, le temps d'obtention est moins que proportionnel au nombre d'entités étudiées, contrairement à ce qui se passe avec les questionnaires ou entretiens. C'est le cas notamment des listes de membres d'organisations (en particulier les conseils d'administration d'entreprises ou les organes dirigeants d'associations), qui ont donné lieu à l'étude de réseaux dits *interlocks* traitant chaque appartenance multiple comme un lien, ainsi que des articles scientifiques, considérés comme une source de réseaux de citations. Le risque, dans ces cas, est que la possibilité d'obtenir assez facilement (par copie d'imprimés, puis plus récemment par récupération de sources nativement numériques) de grandes quantités de données fasse oublier *quel type précis de relation ou d'interaction* on documente finalement (L6). La coprésence dans une institution peut être dans certains cas la trace indirecte de relations interpersonnelles, d'une similitude d'idées, etc., mais elle ne les atteste pas ; la citation d'un article peut être un indice, par exemple, d'une relation d'estime d'un laboratoire à un autre, mais pas toujours. Elle peut reposer aussi sur une controverse ou un conflit scientifique ouvert. Lorsque les données semblent très faciles à construire, le risque est que leur traitement devienne routinier, oublieux des biais des sources et de la nécessité de problématiser. Il est pourtant possible de regarder autrement les listes de noms et les citations (voir notamment Mizruchi, 1996 et Milard, 2011). Et l'ARS peut être renouvelée par l'utilisation de traces de pratiques plus originales – par exemple lorsque le sociologue Baptiste Coulmont (2013) décrit les relations listées ou visibles sur des affiches de pasteurs placardées sur les murs de Paris, pour nourrir une réflexion sur le charisme religieux.

La facilité relative à obtenir des informations sur plusieurs types de liens, ou du moins des liens précisément qualifiés, à partir de traces de pratiques permet, plutôt que de céder à la routine, de construire des données riches et des interprétations subtiles. Souvent, une même source permet non seulement de lister des liens mais aussi de qualifier finement chacun d'entre eux – et parfois chaque entité (L6, L7). C'est le cas dans des sources historiques, mais aussi des sources nativement numériques. Par exemple, une étude récente se fonde sur plus d'un million d'e-mails échangés en 2012 par 193 salarié·es (Brashears et Quintane, 2018). Même si la quantité de données est énorme, le groupe social étudié ne l'est donc pas ; c'est cette focalisation sur des personnes particulières qui permet de collecter et catégoriser certains de leurs attributs et ainsi d'enrichir l'analyse. Alors que les traces numériques sont souvent collectées en masse,

il est donc important de se poser la question : choisir délibérément un échantillon ou une fenêtre plus réduite ne permettrait-il pas de voir, en contrepartie, plus de choses (L3) ? Ainsi, un article pionnier d'analyse des blogs (Lento et alii, 2006) peut comparer la structure de deux liens, de commentaire et d'invitation, analyser leur coévolution, prendre en compte la langue utilisée et ainsi répondre à une question précise, celle de savoir pourquoi certains utilisateurs et trices d'une plate-forme s'arrêtent tandis que d'autres continuent, parce que le choix d'une fenêtre temporelle permettant d'observer « seulement » 4 500 blogs environ a été réfléchi.

L'utilisation de traces, ainsi, même si elle a souvent été privilégiée parce qu'elle semblait économique en temps, peut également conduire à la construction de données riches. En particulier, dans les traces écrites ou numériques, les liens et les entités sont nécessairement qualifiés de manière *emic* (avec les mots propres au terrain : ceux des personnes concernées ou du scribe, du notaire, de la plate-forme) et non pas suivant les catégories de l'analyste. C'est une richesse de ces sources : si l'on prend la peine de conserver ces taxonomies précises, on peut ensuite catégoriser liens et entités de plusieurs manières différentes, selon la problématique (L6, L7). Par exemple, lorsque la source indique le lien de parenté exact entre deux individus, l'analyste peut étudier les effets des « liens familiaux » selon différentes définitions possibles, *emic* ou *etic*, de la famille (Lemerrier, 2006).

Enfin, dans certains cas, on dispose de plusieurs sources différentes pour documenter les mêmes types de liens entre les mêmes entités. Ces cas ont un énorme potentiel de richesse interprétative, à condition que l'on n'ignore pas délibérément ces différences de sources, donc de points de vue sur les liens pour procéder de manière additive (L2). Le procédé additif considère qu'« un lien en provenance de la source A + un lien en provenance de la source B = deux liens », ce qui est toujours tentant, dans la mesure où on a envie de documenter le plus de liens possible. Il peut y avoir de bonnes raisons d'opérer une telle agrégation, à condition qu'on ait bien conscience qu'on obtient alors une vision composite de la réalité, que l'addition d'une troisième source bouleverserait sans doute (en rendant certains individus plus centraux et d'autres moins, en effaçant une frontière, etc.). Mais il est intéressant de se livrer aussi, avant de procéder à l'addition, à une approche plus comparative afin de mieux comprendre les biais de chaque source et de produire ainsi de nouvelles interprétations du phénomène observé. C'est ce qu'a fait, par exemple, Isabelle Rosé (2018), pour comprendre les logiques relationnelles autour d'une reine de France au Moyen Âge, en se demandant notamment si la reine faisait autre chose qu'unir la famille de son père à celle de son mari. De la même façon, Sébastien Dubois et Pierre François (2013) comparent différents discours de poètes contemporains sur les regroupements et influences entre d'autres poètes, pour déterminer quels labels et quelles frontières font consensus et lesquels sont disputés.

Conclusion

Pour ne pas faire des graphes et indicateurs un but, mais un moyen au service d'une compréhension précise et fine des phénomènes relationnels étudiés (l'amitié, le soutien social, l'innovation, les collaborations scientifiques, l'entrepreneuriat, la collégialité, le pouvoir, la formation des marchés, etc.), il faut se souvenir que les données relationnelles sont construites et se poser un certain nombre de questions au fil de leur construction. Que retenir de ce tour d'horizon des sources de données pour l'ARS ?

Aucune méthode de collecte n'est dans l'absolu meilleure qu'une autre ou plus adaptée à l'ARS : tout dépend du terrain et plus encore des questions de recherche, et il est souvent avantageux de combiner au moins deux types de sources.

Une bonne collecte de données prend du temps et ne s'improvise pas. Si l'on veut choisir une fenêtre d'observation adaptée, s'intéresser à plusieurs relations ou interactions bien

définies, documenter des attributs des entités, s'accommoder au mieux des données manquantes, voire prendre en compte des liens négatifs, il faut prévoir une longue enquête et peut-être plus encore une longue préparation d'enquête.

Une banalisation des méthodes de l'ARS et des discussions à leur sujet serait bénéfique tant pour ceux et celles qui la pratiquent déjà que pour ceux et celles qu'elle pourrait intéresser. En effet, bien des éléments que nous avons présentés ici découlent de discussions sur les sources et les modes de collecte qui sont communes aux sciences humaines et sociales en général. Échanger plus largement sur la pratique de l'entretien, l'expérimentation, l'observation et l'analyse de traces écrites et numériques peut contribuer à mieux intégrer l'ARS au sein des sciences humaines et sociales.

De notre expérience et notre vue d'ensemble de formateur et formatrices, il nous semble que dans l'ensemble, les recherches mobilisant l'ARS en France et plus généralement dans le monde francophone prennent en compte la plupart des écueils et arbitrages que nous avons évoqués en première partie. Les différentes formations, écoles d'été du CNRS, conférences, réseaux de recherche (du groupement de recherches « Analyse de réseaux en sciences humaines et sociales », du réseau thématique 26 « Réseaux sociaux » de l'Association française de sociologie, des groupes « Réseaux et histoire » [Res-Hist], « Flux, matrices, réseaux » [FMR], etc.) ont permis une diffusion des compétences d'analyse, mais aussi de collecte de données de réseaux. Cette dernière s'est toutefois principalement faite de manière orale, en petits groupes : elle ne permet pas aisément d'atteindre les nombreux·ses chercheur·ses qui se sont mis à l'ARS depuis les années 2000 dans toutes les disciplines. Pour toucher cet autre public, il faut structurer et mettre par écrit les acquis de ces expériences : cet article constitue un premier pas dans cette direction.

Références

- Baker WE (1984) The social structure of a national securities market. *American Journal of Sociology* 89(4): 775–811.
- Barnes JA (1954) Class and committees in a Norwegian island parish. *Human Relations* 7(1): 39–58.
- Bearman P and Parigi P (2004) Cloning headless frogs and other important matters: Conversation topics and network structure. *Social Forces* 83(2). Oxford Academic: 535–557.
- Bernard HR, Killworth PD and Sailer L (1982) Informant accuracy in social-network data V. An experimental attempt to predict actual communication from recall data. *Social Science Research* 11(1): 30–66.
- Bès M-P (2011) Les chaînes relationnelles entre anciens étudiants : l'usage des carnets d'adresses électroniques. *Réseaux*, 168-169 : 189-214.
- Bidart C, Degenne A and Grossetti M (2011) *La vie en réseau : dynamique des relations sociales*. Paris: Presses universitaires de France.
- Birkett M, Melville J, Janulis P, Phillips G, Contractor N, and Hogan B (2021) Network Canvas: Key decisions in the design of an interviewer-assisted network data collection software suite. *Social Networks* 66: 114–124.
- Blondel V, Guillaume J-L, Lambiotte R and Lefebvre E (2008) Identification multi-échelle de la structure communautaire de très grands graphes. In: *AlgoTel'08*, 2008, p. 61. Available at: <https://hal.inria.fr/inria-00374457> (accessed 5 May 2021).
- Boje DM and Whetten DA (1981) Effects of organizational strategies and contextual constraints on centrality and attributions of influence in interorganizational networks. *Administrative Science Quarterly* 26(3): 378–395.

- Borgatti SP, Mehra A, Brass DJ and Labianca G. (2009) Network analysis in the social sciences. *Science* 323(5916). American Association for the Advancement of Science: 892–895.
- Bourdieu P (1980) Le capital social: notes provisoires. *Actes de la recherche en sciences sociales*, 31(1), 2-3.
- Boullier D (2015) Les sciences sociales face aux traces du big data. Société, opinion ou vibrations ? *Revue française de science politique* 65(5). Presses de Sciences Po: 805–828.
- Boutier J (2001) L'usage historique des archives. In: *Corpus, sources et archives*. Tunis: Institut de recherche sur le Maghreb contemporain, pp. 9–22.
- Boyardjian J (2014) Twitter, un nouveau « baromètre de l'opinion publique » ? *Participations* (8) De Boeck Supérieur: 55-74.
- Brashears ME and Quintane E (2018) The weakness of tie strength. *Social Networks* 55: 104–115.
- Bright D, Brewer R and Morselli C (2021) Using social network analysis to study crime: Navigating the challenges of criminal justice records. *Social Networks* 66: 50–64.
- Burt RS, Marsden PV and Rossi PH (1985) A research agenda for survey network data. *GSS Methodological Report* (39): 1–72.
- Burt RS (2000) The network structure of social capital. *Research in Organizational Behavior*, 22, 345-423.
- Carrington PJ, Scott J and Wasserman S (eds) (2005) *Models and Methods in Social Network Analysis*. Structural Analysis in the Social Sciences. Cambridge: Cambridge University Press.
- Cefaï D (2003) *L'enquête de terrain*. Paris: La Découverte.
- Chauvac N (2016) Création d'entreprises innovantes et relations professionnelles : l'apport de la méthode des narrations quantifiées. *Regards croisés sur l'économie* (19). La Découverte: 76–86.
- Chauvac N, Cloutier L, Defossez A, Akermann G, and de Federico A (2014) Negative Ties, Lost Ties, Latent Ties. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* (121). SAGE Publications Ltd: 5–9.
- Cloutier L (2014) Inventeurs entrepreneurs. In: Chauvin P-M, Grossetti M, and Zalio P-P (eds) *Dictionnaire sociologique de l'entrepreneuriat*. Paris: Presses de Sciences Po, pp. 380–396.
- Coleman JS (1988) Social capital in the creation of human capital. *American Journal of Sociology*, 94, S95-S120.
- Corman SR, Steiner E, Proulx JD, Dutta A, Yahja A, Poole S, Berisha V, Bliss DW (2021) Revisiting the accuracy problem in network analysis using a unique dataset. *Social Networks* 66: 1–9.
- Coulmont B (2013) Tenir le haut de l'affiche : analyse structurale des prétentions au charisme. *Revue française de sociologie* 54(3): 507.
- Coviello N (2005) Integrating qualitative and quantitative techniques in network analysis. *Qualitative Market Research: An International Journal* 8(1): 39–60.
- Crossley N (2008) (Net)Working out: Social capital in a private health club. *The British Journal of Sociology* 59(3): 475–500.
- Crossley N (2010) The Social world of the network. Combining qualitative and quantitative elements in social network analysis. *Sociologica* (1): 1–34.
- Crossley N, Bellotti E, Edwards G, Everett M, Koskinen J, Tranmer M (2015) *Social Network Analysis for Ego-Nets*. London: SAGE Publications Ltd.
- Defossez A (2014) De l'inefficacité du réseau social – Des liens sociaux non mobilisés chez les patients atteints de cancer. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* (121). SAGE Publications Ltd: 44–58.
- Degenne A and Forsé M (2004) *Les réseaux sociaux*. Paris: Armand Colin (1^{re} éd. 1994).

- Desmond M (2012) Disposable ties and the urban poor. *American Journal of Sociology*, 117(5), 1295-1335.
- Dodds PS, Muhamad R and Watts DJ (2003) An experimental study of search in global social networks. *Science* 301(5634). American Association for the Advancement of Science: 827–829.
- Dubois S and François P (2013) Seeing the world through common lenses? The Case of french contemporary poetry. In: Beckert J and Musselin C (eds) *Constructing Quality: The Classification of Goods in Markets*. Oxford: Oxford University Press, pp. 174–193.
- Düring M (2013) *Analysis of Support Networks for Jewish Refugees in the Third Reich*. The Van Leer Jerusalem Institute. Available at: <https://www.youtube.com/watch?v=SIQ7stSU-9w> (accessed 22 April 2020).
- Düring M (2015) From hermeneutics to data to networks: Data extraction and network visualization of historical sources. *The Programming Historian* (4). Available at: <https://programminghistorian.org/en/lessons/creating-network-diagrams-from-historical-sources> (accessed 17 June 2020).
- Eloire F, Penalva-Icher E and Lazega E (2011) Application de l'analyse des réseaux complets à l'échelle interorganisationnelle. *Terrains & Travaux* (19). ENS Paris-Saclay: 77–98.
- Favre G and Brailly J (2015) Salons et définition de normes marchandes : Le cas de la distribution de programmes de télévision en Afrique sub-saharienne. *L'Année sociologique* 65(2). Presses Universitaires de France: 425–456.
- Ferrand A (2014) Les significations des zéros dans les réseaux personnels. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* (121). SAGE Publications Ltd: 10–16.
- Fischer CS (1982) *To Dwell among Friends: Personal Networks in Town and City*. Chicago: University of Chicago Press.
- Fischer CS (2009) The 2004 GSS finding of Shrunken social networks: An Artifact? *American Sociological Review* 74(4). SAGE Publications Inc: 657–669.
- Fischer CS and Bayham L (2019) Mode and interviewer effects in egocentric network research. *Field Methods* 31(3). SAGE Publications Inc: 195–213.
- Fitzhugh SM and Butts CT (2018) Patterns of co-membership: Techniques for identifying subgraph composition. *Social Networks* 55: 1–10.
- Freeman LC (2004) *The Development of Social Network Analysis. A Study in the Sociology of Science*. Vancouver: Empirical Press.
- Godechot O (2012) Réseaux et classes sociales : le malentendu ? *Revue Française de Socio-Économie* (10). La Découverte: 265–271.
- Gourdon V (2008) Les témoins de mariage civil dans les villes européennes du XIX^e siècle : quel intérêt pour l'analyse des réseaux familiaux et sociaux ? *Histoire, économie & société* 27(2): 61–87.
- Granovetter MS (1974) *Getting a Job. A Study of Contacts and Careers*. Cambridge: Harvard University Press.
- Granovetter MS (2003) Ignorance, knowledge, and outcomes in a small world. *Science* 301(5634). American Association for the Advancement of Science: 773–774.
- Grossetti M (2011) Les narrations quantifiées. Une méthode mixte pour étudier des processus sociaux. *Terrains & Travaux* (19). ENS Paris-Saclay: 161–182.
- Grossetti M and Bès M-P (2001) Encastrements et découplages dans les relations science-industrie. *Revue française de sociologie* 42(2): 327–355.
- Hagman EP (1933) The companionship of preschool children. In: Stoddard GD (ed.) *University of Iowa Studies in Child Welfare*. Iowa City: University of Iowa, pp. 10–69.
- Hällsten M, Edling C and Rydgren J (2015) The effects of specific occupations in position generator measures of social capital. *Social Networks* 40: 55–63.

- Harrigan NM, Labianca GJ and Agneessens F (2020) Negative ties and signed graphs research: Stimulating research on dissociative forces in social networks. *Social Networks* 60: 1–10.
- Holland PW and Leinhardt S (1973) The structural implications of measurement error in sociometry. *The Journal of Mathematical Sociology* 3(1). Routledge: 85–111.
- Jourdan M (2019) Les rencontres des Irlandais hors d'Irlande au VIIe siècle. Comprendre et dépasser les biais des sources par l'analyse de réseaux. *Hypothèses* 22(1). Éditions de la Sorbonne: 263–277.
- Killworth PD, McCarty C, Bernard HR and House M (2006) The accuracy of small world chains in social networks. *Social Networks* 28(1): 85–96.
- Kitts JA and Leal DF (2021) What is(n't) a friend? Dimensions of the friendship concept among adolescents. *Social Networks* 66: 161–170.
- Kleinfield JS (2002) Could it be a big world after all? The 'six degrees of separation' myth. *Society* (39): 61–66.
- Knox H, Savage M & Harvey P (2006) Social networks and the study of relations: networks as method, metaphor and form. *Economy and society*, 35(1), 113-140.
- Korte C and Milgram S (1970) Acquaintance networks between racial groups: Application of the small world method. *Journal of Personality and Social Psychology* 15(2). US: American Psychological Association: 101–108.
- Kros M, Jaspers E and van Zalk M (2021) Avoidance, antipathy, and aggression: A three-wave longitudinal network study on negative networks, status, and heteromisos. *Social Networks* 64: 122–133.
- Laumann EO, Marsden PV and Prensky D (1983) The boundary specification problem in network analysis. In: Burt RS and Minor MJ (eds) *Applied Network Analysis*. Beverly Hills: Sage, pp. 18–34.
- Lazega E (2014) *Réseaux sociaux et structures relationnelles*. Paris: PUF (1^{re} éd. 2007).
- Lemerrier C (2006) Liens privés et régulation de l'économie: la famille et l'institution (Paris, XIXe siècle). *Revue d'histoire du XIXe siècle*, 33(2): 23–53.
- Lemerrier C (2015) Taking time seriously: How do we deal with change in historical networks? In: Düring M, Gamper M, and Reschke L (eds) *Knoten Und Kanten III. Soziale Netzwerkanalyse in Politik- Und Geschichtswissenschaft*. Bielefeld: Transcript Verlag, pp. 183–211.
- Lemerrier C (forthcoming) Historical and archaeological network data. In: Brughmans T, Mills BJ, Munson JL, et al. (eds) *Oxford Handbook of Archaeological Network Research*. Oxford: Oxford University Press.
- Lento T, Welser HT, Gu L and Smith M(2006) The ties that blog: Examining the relationship between social ties and continued participation in the wallop weblogging system. *3rd annual workshop on the Weblogging Ecosystem: aggregation, Analysis and Dynamics, Edimburgh*. Available at: <https://www.connectedaction.net/wp-content/uploads/2009/08/2006-WWW-Lento-Welser-Gu-Smith-TiesThatBlog.pdf>.
- Lin N, Dayton PW and Greenwald P (1978) Analyzing the instrumental use of relations in the context of social structure. *Sociological Methods & Research* 7(2). SAGE Publications Inc: 149–166.
- Lin N, Fu Y and Hsung R-M (2001) The position generator: Measurement techniques for investigations of social capital. In: Lin N, Cook K, and Burt RS (eds) *Social Capital. Theory and Research*. New York: Routledge, pp. 57–81.
- Lin N (2002) *Social capital: A theory of social structure and action* (No. 19). Cambridge university press.
- Marin A (2004) Are respondents more likely to list alters with certain characteristics? Implications for name generator data. *Social Networks* 26(4): 289–307.

- Marsden PV (2003) Interviewer effects in measuring network size using a single name generator. *Social Networks* 25(1): 1–16.
- Marineau JE and Labianca GJ (2021) Positive and negative tie perceptual accuracy: Pollyanna principle vs. negative asymmetry explanations. *Social Networks* 64: 83–98.
- Marsden PV (2011) Survey methods for network data. *The SAGE handbook of social network analysis*, 25, 370-388.
- McCallister L and Fischer CS (1978) A Procedure for surveying personal networks. *Sociological Methods & Research* 7(2). SAGE Publications Inc: 131–148.
- McCarty C, Lubbers MJ, Vacca R and Molina JL (2019) *Conducting Personal Network Research: A Practical Guide*. New York: Guilford Publications.
- McPherson M, Smith-Lovin L and Brashears ME (2006) Social isolation in america: changes in core discussion networks over two decades: *American Sociological Review* 71(3). SAGE PublicationsSage CA: Los Angeles, CA: 353–375.
- Mercklé P (2013) La « découverte » des réseaux sociaux. À propos de John A. Barnes et d'une expérience de traduction collaborative ouverte en sciences sociales. *Réseaux* (182). La Découverte: 187–208.
- Mercklé P (2016) *Sociologie des réseaux sociaux*. Paris: La Découverte (1^{re} éd. 2004).
- Milard B (2011) Dynamiques relationnelles d'un article scientifique : « Roger et al. (2004) » et ses réseaux. *Terrains & Travaux* (21). ENS Paris-Saclay: 141–160.
- Milgram S (1967) The small-world problem. *Psychology Today* (1): 61–67.
- Mische A (2003) Cross-Talk in movements: Rethinking the culture-network link. In Diani M and McAdam D. *Social Movements and Networks: Relational Approaches to Collective Action*, Oxford University Press, pp. 258-280.
- Mizruchi MS (1996) What do interlocks do? An analysis, critique, and assessment of research on interlocking directorates. *Annual Review of Sociology* 22: 271–298.
- Moreno JL (1953) *Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Sociodrama*. Beacon: Beacon House.
- Offer S and Fischer CS (2018) Difficult people: Who is perceived to be demanding in personal networks and why are they there? *American Sociological Review* 83(1). SAGE Publications Inc: 111–142.
- Osa M (2003) *Solidarity and Contention: Networks of Polish Opposition*. Minneapolis Minn.: University of Minnesota Press.
- Padgett JF and Ansell CK (1993) Robust action and the rise of the Medici, 1400-1434. *American Journal of Sociology* 98(6): 1259–1319.
- Paik A and Sanchagrin K (2013) Social isolation in America: An artifact. *American Sociological Review* 78(3). SAGE Publications Inc: 339–360.
- Pál J, Stadtfeld C, Grow A and Takács K (2016) Status perceptions matter: Understanding disliking among adolescents. *Journal of Research on Adolescence* 26(4): 805–818.
- Paugam S (ed.) (2012) *L'enquête sociologique*. Paris: PUF.
- Perry BL, Pescosolido BA and Borgatti SP (2018) *Egocentric Network Analysis: Foundations, Methods, and Models*. Structural Analysis in the Social Sciences. Cambridge: Cambridge University Press.
- Putnam RD (2000) Bowling alone: America's declining social capital. In *Culture and Politics* (pp. 223-234). Palgrave Macmillan, New York.
- RESTIC 2017 (n.d.), In: *Structurations des Mondes Sociaux*. Available at: <https://sms.univ-tlse2.fr/tic-et-reseaux-sociaux/restic-2017-483214.kjsp?RH=1350906327519> (accessed 17 June 2020).
- Robins G (2015) *Doing Social Network Research. Network-Based Research Design for Social Scientists*. London: SAGE Publications Ltd.

- Roda R (2014) On negative genesis – The role of dyadic and network attributes in the emergence of hostile relationships. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* (123). SAGE Publications Ltd: 20–40.
- Rosé I (2018) Autour de la reine Emma (vers 890-934). Réseaux, itinéraire biographique féminin et questions documentaires au début du Moyen Âge central. *Annales HSS* 73(4). Éditions de l'EHESS: 817–847.
- Ruiz É (2011) À propos des sources. In: *Devenir historien-ne*. Available at: <https://devhist.hypotheses.org/849> (accessed 12 June 2020).
- Scott J (2000) *Social Network Analysis*. London: SAGE.
- Scott J and Carrington PJ (eds) (2011) *The SAGE Handbook of Social Network Analysis*. London-Thousand Oaks: SAGE.
- Shipilov A and Gawer A (2020) Integrating research on interorganizational networks and ecosystems. *Academy of Management Annals* 14(1): 92–121.
- Singh J, Hansen MT and Podolny JM (2010) The World Is Not Small for Everyone: Inequity in Searching for Knowledge in Organizations. *Management Science* 56(9). INFORMS: 1415–1438.
- Small ML (2013) Weak ties and the core discussion network: Why people regularly discuss important matters with unimportant alters. *Social networks*, 35(3), 470-483.
- Small ML (2017) *Someone To Talk To*. Oxford: Oxford University Press.
- Stehlé J, Voirin N, Barrat A, Cattuto C, Isella L, Pinton JF, Quaghiotto M, Van den Broeck W, Régis C, Lina B and Vanhems P (2011) High-Resolution measurements of face-to-face contact patterns in a primary school. *PLoS ONE* 6(8).
- Stevenson W, Davidson B, Manev I and Walsh K (1997) The small world of the university: A classroom exercise in the study of networks. *Connections* 20(2): 23–33.
- Stoltenberg D, Maier D and Waldherr A (2019) Community detection in civil society online networks: Theoretical guide and empirical assessment. *Social Networks* 59: 120–133.
- Travers J and Milgram S (1969) An experimental study of the small world problem. *Sociometry* 32(4): 425–443.
- Van Der Gaag M and Snijders TAB (2005) The Resource Generator: social capital quantification with concrete items. *Social Networks* 27(1): 1–29.
- Wasserman S and Faust K (1994) *Social Network Analysis. Methods and Applications*. Cambridge: Cambridge University Press.
- White HC, Boorman SA and Breiger RL (1976) Social structure from multiple networks. I. Blockmodels of roles and positions. *American Journal of Sociology* 81(4): 730–780.
- Williams T (2005) Cooperation by design: structure and cooperation in interorganizational networks. *Journal of Business Research* 58(2): 223–231.