

**Recueil du Symposium de 2022 de Statistique Canada :  
Désagrégation des données : dresser un portrait de données plus représentatif  
de la société**

**De la théorie à la pratique : Quelques  
enseignements de la mise en œuvre de la  
méthode « Network Sampling with Memory »**

par Aurélie Santos et Géraldine Charrance

Date de diffusion : le 25 mars 2024



Statistique  
Canada

Statistics  
Canada

Canada

## **De la théorie à la pratique : Quelques enseignements de la mise en œuvre de la méthode « Network Sampling with Memory »**

Aurélié Santos et Géraldine Charrance<sup>1</sup>

### **Résumé**

Pour pallier les défauts classiques des méthodes par chaînage, la méthode d'échantillonnage appelée « Network sampling with memory » a été développée. Sa particularité est de recréer, au fil du terrain, une base de sondage de la population cible composée des personnes citées par les répondants et de tirer aléatoirement les futurs enquêtés dans cette base, minimisant ainsi les biais de sélection. Expérimentée pour la première fois en France entre septembre 2020 et juin 2021 dans le cadre d'une enquête auprès des immigrés chinois en Ile-de-France (ChIPRe), cette présentation est une restitution des difficultés rencontrées lors de la collecte, parfois contextuelles, en raison de la pandémie, mais pour la plupart inhérentes à la méthode.

Mots Clés : Échantillonnage ; collecte ; population difficile à joindre.

### **1. NSM, une méthode innovante et ambitieuse pour l'étude des populations cachées**

#### **1.1 Les immigrés chinois en Ile-de-France, une population difficile à enquêter**

Schiltz situe l'émergence dans la statistique publique des enquêtes sociales portant sur des populations dites « marginales » aux années 1990 (Schiltz, 2005). Le besoin en données est important pour ces populations « en marge », souvent en situation de précarité ou exposées à des risques sociaux ou sanitaires. Cependant, ces populations sont difficiles à enquêter car elles sont de petite taille, ses membres ne sont pas identifiables dans une base de sondage et peuvent même vouloir cacher leur appartenance à la population d'intérêt pour éviter d'être stigmatisés (Marpsat & Razafindratsima, 2010). Le recours à des enquêtes en population générale pour capter ces populations ne permet généralement pas d'avoir un échantillon suffisamment grand pour mener des analyses statistiques. Ces contraintes excluent donc le recours à des enquêtes probabilistes en population générale pour capter et caractériser ces populations.

Les immigrants chinois en Île-de-France s'inscrivent dans cette catégorie des populations difficiles à atteindre : dans l'enquête Trajectoires et Origines (Beauchemin et al. 2016) menée en 2008, on ne compte que 69 répondants nés en Chine. D'autre part, les sources de données comme le recensement français, l'échantillon démographique permanent (EDP) ou Fidéli contiennent peu d'informations pour caractériser ces groupes de populations immigrées et peuvent souffrir de défaut de couverture (manquer les migrants en situation irrégulière).

#### **1.2 Les méthodes d'enquêtes par réseau particulièrement adaptées à ce type de populations**

Les méthodes de sondage par chaînage consistent à enquêter au sein de réseaux en mettant à contribution les répondants pour en recruter de nouveaux. Elles ne nécessitent donc pas de base de sondage. Le recours à ces méthodes est motivé également par d'autres raisons : une croyance selon laquelle les répondants sont plus susceptibles de participer lorsqu'ils sont invités à répondre par des connaissances plutôt que contactés au hasard ; et un désir de comprendre la structure des réseaux et le rôle des relations sociales.

---

<sup>1</sup>Aurélié Santos, Institut national d'études démographiques, 9 cours des Humanités, CS 50004, 93322 Aubervilliers, France (aurelie.santos@ined.fr); Géraldine Charrance, Institut national d'études démographiques, 9 cours des Humanités, CS 50004, 93322 Aubervilliers, France (geraldine.charrance@ined.fr).

Ces méthodes ont longtemps souffert de la réputation de produire des estimations biaisées. Cependant, des études récentes basées sur la méthode « Respondent-driven sampling » (RDS) ont montré que sous certaines hypothèses, la probabilité d'être échantillonné pouvait être estimée par le nombre de liens que chaque répondant a avec les autres membres de la population. Les hypothèses sont: (1) la population est formée d'un seul réseau : tous les individus sont connectés entre eux, plus ou moins directement (hypothèse essentielle permettant de mobiliser la théorie de la chaîne de Markov et ainsi argumenter en faveur de la convergence des estimateurs vers des estimateurs sans biais). (2) Les liens entre les individus sont réciproques. (3) Les individus recrutent aléatoirement parmi leurs connaissances. Malgré cela, la précision de la méthode reste problématique : alors que la variance d'échantillonnage d'un échantillon aléatoire simple est inversement proportionnelle à la taille de l'échantillon, la précision des RDS est fonction à la fois de la taille de l'échantillon et de la structure du réseau.

L'objectif de cette communication est donc de mobiliser la méthode « Network sampling with memory » (NSM). L'approche de marche aléatoire de base est améliorée grâce à l'incorporation d'informations sur la topographie locale du réseau et sur les cas récemment échantillonnés pour gagner en efficacité lors de l'échantillonnage (Avin et Krishnamachari 2008). La méthode NSM est supposée améliorer l'efficacité de l'échantillonnage à partir d'un réseau en recueillant des données sur le réseau auprès des répondants, qui sont utilisées pour révéler progressivement la liste des membres de la population. Au fur et à mesure, la liste des personnes désignées dans l'enquête tend à ressembler à la liste complète des membres de la population, ce qui permet théoriquement au NSM de se rapprocher du sondage aléatoire simple.

### **1.3 Les apports de la variante Network Sampling with Memory (NSM)**

Les concepteurs de la méthode Network sampling with memory ont voulu développer une méthode asymptotiquement sans biais (comme l'approche RDS lorsque ses hypothèses sont satisfaites) et dont la variance d'échantillonnage et donc l'erreur absolue moyenne diminuent rapidement avec la taille de l'échantillon collecté. Pour cela, la méthode NSM recourt à deux modes d'échantillonnage fonctionnant en tandem pour améliorer la précision des estimations : un mode « List » qui garantit des estimations asymptotiques sans biais (Naive list mode et Even sampling) et un mode « Search ». Le processus d'échantillonnage est utilisé pour découvrir la liste des membres de la population (Search), puis l'échantillonnage se fait dans cette liste avec remplacement (List). Chaque répondant se voit, en fin de questionnaire, demander de fournir les coordonnées de ses amis qui sont membres de la population A (cette liste est ici désignée par le terme « roster »). À chaque étape, de nouveaux membres du réseau sont ajoutés à la liste (roster total), qui comprend tous les membres de la population désignés par les répondants. Les modes interviennent de manière successive au fil de la collecte : d'abord le Naive list, puis le Search, et enfin l'Even sampling. Afin de déterminer le mode à mobiliser, on s'appuie sur des indicateurs révélant le niveau d'exploration du réseau.

Naive list mode : Démarrer la collecte aléatoirement

La collecte démarre par l'identification de graines, qui sont les premiers enquêtés. Leurs réponses génèrent un premier réseau. En début de collecte, on tire aléatoirement un individu parmi les premiers identifiés dans le réseau (appelés « nœuds »). Ayant une connaissance très faible du réseau à ce moment-là, il convient de mobiliser la technique de sondage la plus simple et égalitaire : le sondage aléatoire simple.

Search mode : Assurer l'exploration du réseau

L'objectif est d'enquêter des pans du réseau encore inexplorés. Pour cela, il faut repérer les répondants les plus susceptibles de nous y conduire : ces individus sont appelés « nœuds-ponts » (ponts entre sous-parties du réseau). On calcule, pour chaque répondant, sa probabilité d'être un nœud-pont, fondée sur la proportion de personnes citées une seule fois parmi leurs « amis ». Après identification des 5 individus les plus susceptibles d'être des nœuds-ponts, on en sélectionne un (proportionnellement à sa probabilité d'être un nœud-pont), puis on tire aléatoirement un de ses amis parmi ceux cités une seule fois et non enquêtés. Il s'agit d'un tirage à deux degrés.

Even sampling mode (List mode) : Homogénéiser les expositions au tirage

Pour atténuer l'effet du sur-échantillonnage des nœuds apparus tôt, le mode Even sampling permet d'homogénéiser les CSR (cumulative sampling rate). À chaque tirage, quel que soit le mode, le CSR des individus exposés à l'échantillonnage est incrémenté de leur probabilité d'être sélectionné. Actualisé dès lors que l'individu est candidat à l'échantillonnage, le CSR mesure l'exposition des individus aux tirages successifs. Homogénéiser les CSR consiste donc à exposer au tirage les individus nouveaux ou jamais soumis au tirage jusqu'alors. Pour cela, on exclut du tirage

les nœuds les plus exposés au tirage précédemment, et ayant un CSR important. Ici, on recourt à un tirage aléatoire simple parmi les nœuds ayant un CSR inférieur à l'ESR (Even sampling rate) (ou les 100 CSR les plus bas si le volume de nœuds remplissant la condition est inférieur à 100).

La sortie du mode Search repose sur trois critères : (1) avoir collecté plus de 50 questionnaires, (2) avoir une taille de réseau supérieure à 200 ou si les 5 dernières interviews n'ont pas permis de découvrir de nouveaux individus et enfin (3) lorsque le paramètre P1 (nombre d'individus cités une seule fois et non enquêtés) passe en dessous du seuil A1 (fixé par les concepteurs de la méthode). Lorsque P1 est bas, cela signifie que beaucoup d'individus ont été cités plusieurs fois, c'est-à-dire que l'on a atteint un certain niveau d'exploration. Il convient alors de passer en Even sampling afin de soumettre aux tirages les individus jusqu'alors ignorés, et d'équilibrer les CSR. En pratique, les deux premiers critères ont été assez rapidement atteints. C'est le critère  $P1 < A1$  qui nous est apparu comme le plus discriminant, celui qui nous a contraint à rester en mode Search jusqu'à la fin de la collecte.

## **2. Passer de la théorie à la pratique : la complexe mise en œuvre de la méthode NSM**

La collecte des données de l'enquête auprès des immigrants chinois en Ile-de-France (ChIPRe) s'est déroulée de septembre 2020 à juin 2021 et a mobilisé 10 enquêteur·rices et une équipe de 5 gestionnaires de collecte (informaticien, statisticien·nes, ingénieur·e d'études). La partie suivante aborde notre expérience du passage en production de NSM. En dépit d'une intense préparation et d'une automatisation conséquente des processus de collecte (algorithmes de matching et de tirage, application web de collecte et de suivi conçue ad-hoc), de nombreux ajustements en cours de collecte ont été nécessaires.

### **2.1 Finaliser le protocole d'échantillonnage et de collecte « sur le tas »**

#### **2.1.1 Une élaboration des principes de tirage par « tâtonnement »**

La première difficulté à laquelle s'est trouvée confrontée l'équipe gestionnaire a été l'élaboration de règles encadrant les tirages. La méthode implique de suivre le dévoilement progressif d'un réseau, et donc nécessite de renouveler régulièrement les tirages afin d'alimenter la liste des individus « à enquêter » et de progresser au sein de ce réseau. La méthode ayant très peu été testée, nous ne disposions pas de recommandations sur les fréquence et taille optimales des tirages. Concernant la fréquence, nous avons commencé par produire de très petits tirages (entre 5 et 10 individus) deux fois par semaine. Cette stratégie a permis de suivre de près l'évolution du réseau au démarrage de la collecte mais s'est rapidement avérée trop chronophage. Nous avons alors décidé de ne réaliser qu'un seul tirage par semaine.

Pour déterminer le nombre d'individus tirés à chaque tirage, nous mobilisons les paramètres suivants : la charge d'enquête restante pour chaque enquêteur·rice ; leur charge de relance et d'identification des rosters incomplets; la taille du réseau au moment du tirage par rapport à sa taille lors du précédent. L'inconvénient est qu'il faut d'abord identifier les doublons afin de connaître le nombre de nouveaux individus ajoutés dans le roster total. Après diverses expérimentations, nous avons décidé de produire des tirages ne dépassant jamais la moitié des nouveaux arrivés dans le réseau depuis le précédent tirage

Une fois le tirage réalisé, reste à répartir les échantillonnés entre enquêteur·rices. Nous avons démarré le terrain avec un objectif de respect de l'équité entre enquêteurs : donner à chacun le même nombre de nouveaux cas pour égaliser les chances d'avancer dans le réseau. Ce choix s'est rapidement montré intenable car en contradiction avec le respect des « filiations ». Les enquêteur·rices souhaitaient suivre « leur » réseau, c'est-à-dire enquêter les cités de « leurs » enquêtées (si échantillonnés). Or, une répartition équitable d'échantillonnés entre eux impliquait des « transferts de filiation » beaucoup trop coûteux (il faut négocier avec un répondant que son réseau soit enquêté par un autre enquêteur inconnu !). Nous avons donc opté pour un principe de strict suivi des « filiations » au moment de la répartition des échantillonnés. Avec ces diverses contraintes, il est possible de programmer un nombre total d'échantillonnés, mais il est impossible d'anticiper la charge nouvelle que recevra chaque enquêteur·rice à chaque tirage. Cela induit une forte complexité dans le management de l'équipe sur le terrain.

#### **2.1.2 Des modifications de l'algorithme d'échantillonnage en cours de terrain**

La mise en production de l'algorithme en collecte réelle s'est accompagnée de diverses surprises que des simulations sur des réseaux fictifs ne permettaient pas d'anticiper. Nous avons adapté les critères de calcul permettant de déterminer la liste des rosters éligibles à chaque tirage en mode Search, en excluant les nœuds déjà échantillonnés auparavant afin, entre autre, de limiter la surexposition artificielle au tirage des rosters déjà candidats au tirage précédemment.

Face au rythme très lent d'avancée dans le réseau et au regard de nos objectifs et de la durée de l'enquête, nous avons décidé d'augmenter le seuil A1 (cf. partie 1) afin d'avoir plus de chances de l'atteindre et d'expérimenter la « sortie » du mode Search. Ce seuil n'a pas été décidé arbitrairement, il a été tiré de l'article fondateur de la méthode NSM, dans lequel diverses versions de ce seuil sont présentées (Mouw, 2012).

### **2.1.3 Des adaptations du protocole face aux réalités du terrain**

Face à la lente progression dans le réseau, nous avons dû prolonger la durée de l'enquête, qui devait initialement prendre fin en mars 2021, jusqu'en juin 2021. Cette prolongation s'est accompagnée d'un renouvellement partiel de l'équipe des enquêteur·rices et donc d'une nouvelle formation en cours d'enquête.

En raison des difficultés d'accès à certains groupes sociaux et régionaux, et pour compenser l'extinction progressive de certaines branches du réseau, de nouvelles graines ont été recrutées et ajoutées en cours de collecte. Les personnes originaires de la région de Wenzhou, les étudiants, et les personnes les plus précaires ont fait l'objet d'une attention particulière à cet égard. La réalité du terrain nous a rapidement imposé plus de souplesse dans l'attribution des chèques cadeaux pour remercier les enquêté·es. Le protocole prévoyait un chèque de 15€ pour le questionnaire et de 20€ pour le remplissage d'un roster avec 6 contacts renseignés. Il nous a semblé plus réaliste de distribuer le chèque roster à partir de 3 contacts cités.

Finalement, la mise en œuvre de la méthode NSM, sans guide pratique existant pour son passage en production, a induit de nombreux tâtonnements pour cerner le fonctionnement de l'algorithme, pour concevoir une stratégie de tirages proportionnés, un suivi des graines adapté et des protocoles plus réalistes.

## **2.2 Découvrir un algorithme aux effets contre-intuitifs et parfois désincitatifs**

### **2.2.1 La préférence paradoxale pour des rosters petits et sans doublon**

Tandis que nos consignes aux enquêteur·rices insistaient sur l'importance de collecter de grands rosters, nous avons peu à peu réalisé que l'algorithme favorisait à l'inverse les petits rosters lors des tirages. Selon notre interprétation, cet « effet pervers » est dû au fait que l'algorithme considère les petits rosters comme décrivant probablement des réseaux de personnes plus isolées, ou de potentiels nœuds-ponts vers d'autres sous-réseaux plus difficiles d'accès.

Autre découverte : l'algorithme défavorise de surcroît les rosters contenant des doublons (un doublon est un individu déjà cité au préalable par une autre enquêtée). La quête des doublons fait toutefois partie des objectifs de l'enquête : plus on a de doublons, plus l'indicateur de saturation du réseau baisse (P1, cf. partie 1) et plus on a de chances de passer sous le seuil de sortie du mode Search (A1) pour basculer en Even sampling. Les enquêteur·rices sont donc formé·es à collecter suffisamment d'informations pour rendre l'identification des doublons réalisable grâce à l'algorithme de matching conçu spécialement pour l'enquête.

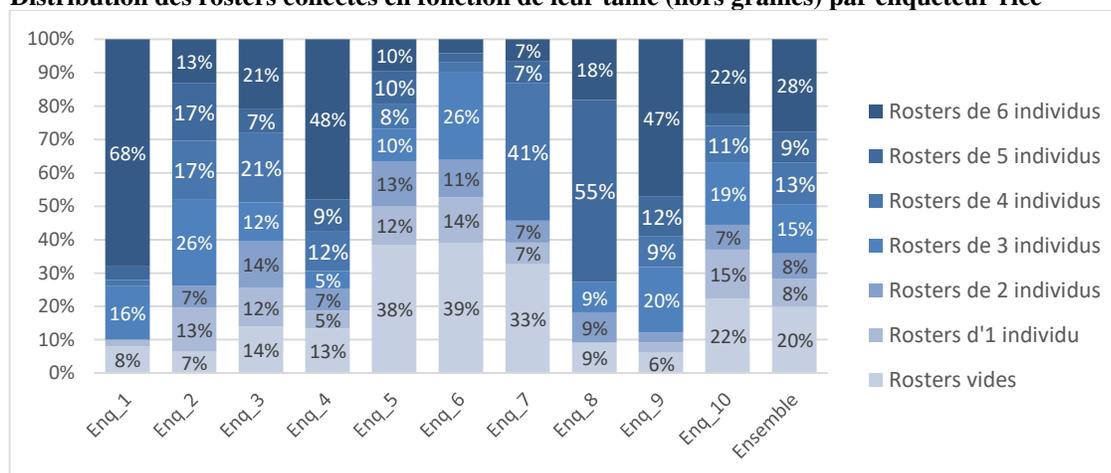
### **2.2.2 ... et induisant des stratégies et des biais enquêteurs non négligeables**

Ces tendances de l'algorithme sont contre-intuitives et contre-productives car les enquêteur·rices finissent par s'en rendre compte et par adapter leur comportement pour minimiser les efforts de collecte. Une enquêtrice a rapidement compris ces deux effets et s'est mise à collecter de petits rosters contenant de 1 à 3 citations, s'assurant ainsi de limiter la présence de doublons et de maximiser ses chances de voir ses rosters sélectionnés au tirage suivant (figure 2.2.2-1, enq 6). Elle était donc certaine d'avoir une liste de « à enquêter » toujours remplie, tout en fournissant le minimum d'effort de négociation des rosters. Or l'avancée dans le réseau, et à terme le passage d'un mode d'échantillonnage à l'autre, repose sur un effort collectif visant à dévoiler le maximum de liens dans le réseau enquêté. De plus, la structure du réseau en ressort biaisée : ses enquêtées disposaient sans doute de réseaux plus étendus que ce dont leur roster

témoignait en définitive, et ses enquêtées au roster vide avaient probablement un réseau, plus difficile à négocier, mais pas inexistant.

**Figure 2.2.2-1**

**Distribution des rosters collectés en fonction de leur taille (hors graines) par enquêteur-riche**



Sans aller jusqu'à ce niveau de rationalisation, le biais « anti-doublon » reste désincitatif : la présence d'un doublon n'est pas perçue comme une bonne nouvelle par les enquêteur-rices car elle implique que le roster ne sera pas prioritaire au tirage. Le travail de collecte des informations identifiantes (les plus difficiles à négocier) qui permet la détection du doublon n'est donc pas valorisé. L'algorithme n'incite pas les enquêteur-rices à « produire » les doublons.

### 2.2.3 Une temporalité d'échantillonnage distante de la réalité du terrain

Enfin, la temporalité de l'algorithme n'est pas la même que celle du terrain. Certains rosters collectés en décembre ont pu être éligibles à l'échantillonnage en mars, demandant un travail de négociation difficile sur le terrain pour les enquêteur-rices.

Ces résultats de l'échantillonnage sont d'autant moins acceptés par les enquêteur-rices qu'ils se cumulent à d'autres, décrits précédemment : si l'algorithme mobilise des rosters anciens moins exploités, c'est aussi parce que l'offre de rosters en cours est moins propice à être échantillonnée, en raison notamment des doublons. Au découragement lié à la non sélection des rosters de taille conséquente avec doublons bien identifiés, s'ajoute la difficulté d'aller négocier la poursuite de la collecte auprès d'enquêtées sollicitées plusieurs mois en amont.

Pour conclure, l'algorithme produit des effets contre-intuitifs rendant la gestion de collecte ardue. Un effort de pédagogie constant de la part des équipes gestionnaires a été nécessaire pour rendre le processus de collecte acceptable et, autant que possible, cohérent avec les principes de la méthode NSM. Enfin, la méthode nécessite un cadre de passation en face-à-face, ce dont nous avons été privés par la pandémie de Covid-19. La suspension du face-à-face induite par le second confinement et la reprise de l'épidémie à l'automne ont eu des impacts conséquents sur la collecte.

## 2.3 Expérimenter un contexte sanitaire très défavorable aux enquêtes en face-à-face

### 2.3.1 Une population d'autant plus difficile à joindre dans ce contexte de crise

Les immigrés chinois en Île-de-France ont très tôt montré une grande méfiance liée aux risques sanitaires, étant informés du déroulé de l'épidémie en Chine. Les retours du terrain ont montré qu'une partie d'entre eux avaient d'autres préoccupations en tête, après avoir essuyé de lourdes pertes avec la fermeture de commerces et de restaurants, ou parce qu'il fallait maintenir un lien avec les membres de la famille en Chine, souvent eux-mêmes en difficulté à cause de l'épidémie.

Les enquêteur·rices ont aussi témoigné d'un effet de contraction du réseau de sociabilité : les enquêté·es ne fréquentent plus autant leurs amis et connaissances qu'avant la crise, et il est d'autant moins approprié qu'ils les sollicitent à l'occasion d'une enquête par questionnaire.

Certaines sous-populations se sont montrées particulièrement inaccessibles en distanciel. Les Wenzhou semblent appartenir à des réseaux très fermés, peu disponibles et peu intéressés par la problématique de recherche mise en avant. Les immigrés sans-papiers constituent une sous-population vulnérable et peu prompte à citer des connaissances, surtout à distance. Ils vivent souvent dans des logements partagés où l'intimité fait défaut pour répondre correctement à l'enquête par téléphone. Une forte implication sur le terrain pour tisser un lien de confiance avec ces enquêté·es a été nécessaire en début de collecte et l'arrêt du face-à-face n'a pas permis d'explorer ce réseau autant qu'attendu.

### **2.3.2 Des leviers de négociation largement altérés par l'abandon du face-à-face**

Collecter des lignes de roster repose sur la confiance tissée lors de la passation du questionnaire, et la phase de négociation pour obtenir les coordonnées des amis des enquêté·es est beaucoup plus simple à aborder en face-à-face qu'en distanciel. L'enquêteur·rice peut demander à l'enquêté·e d'appeler ses amis dans l'immédiat et l'accompagner dans l'argumentaire. Si ce « contrôle » sur la négociation de l'enquêté·e avec ses amis est envisageable en face-à-face, il ne l'est plus du tout en distanciel. En conséquence, obtenir un roster complet demande à l'enquêteur·rice de relancer l'enquêté·e ultérieurement et tout le rythme de collecte s'en trouve ralenti. Enfin, ne pas pouvoir remettre le chèque cadeau immédiatement en main propre à l'issue de la passation a rendu l'incitation financière moins efficace et la recommandation auprès des pairs moins motivée.

La pandémie a complexifié notre enquête auprès d'une population déjà difficile d'accès au départ et éprouvée par les stigmates d'un virus identifié comme provenant de son pays natal. La collecte des rosters a été impactée par la faiblesse des liens relationnels entretenus à cette période et par l'impossibilité de négocier en face-à-face, chèque-cadeau et autres stratégies à l'appui. Enfin, le mode distanciel est peu adapté à ce type de protocole reposant essentiellement sur la recommandation par les pairs, et donc la confiance et les liens établis sur le terrain.

## **3. Conclusion**

Malgré ces nombreuses difficultés, 501 questionnaires ont été collectés, conduisant à un taux de réponse de 60 % (844 individus échantillonnés). En plus des questionnaires, 1689 « liens » ont été recensés via les rosters, soit 3,37 individus cités par interview (sur un maximum de 6 citations), ce qui représente 1 523 individus distincts dans le réseau. Nous n'avons jamais rempli les conditions pour sortir du mode Search. L'indicateur P1 (proportion d'individus non enquêtés et cités une seule fois) n'est jamais descendu en dessous de 0,6 alors que nous avons fixé le seuil A1 à 0,4. Nous nous sommes posés des questions sur les conséquences du non-passage en mode Even sampling sur la qualité et la représentativité de l'échantillon, et sur la validité de la méthode dans ce cas de figure.

Il est difficile de prédire ce qu'auraient été nos résultats si notre collecte n'avait pas été affectée par le contexte sanitaire qui a eu un impact fort sur l'acceptabilité de l'enquête, sur la recommandation par les pairs et sur le rythme de collecte. Toutefois, certaines difficultés semblent intrinsèquement liées à la méthode NSM. L'imprévisibilité de la « charge enquêteur », la préférence pour les petits rosters et l'aversion pour les doublons sont des effets qui vont à l'encontre de l'objectif initial de collecter des rosters de taille conséquente pour faire avancer l'enquête, et qui altèrent l'acceptabilité de la méthode par ceux qui doivent l'éprouver au quotidien. NSM apparaît ainsi comme une méthode conçue à partir de réseaux théoriques ou déjà dévoilés mais qui peine à mesurer le coût pratique du processus même de dévoilement du réseau. Notre article constitue un apport en ce sens : il permet un éclairage concret sur la mise en application d'une telle méthodologie. En définitive, la méthode nous semble prometteuse, mais sans doute plus adaptée pour une expérimentation auprès d'une population mieux connue et de taille plus modeste. Cela permettrait de mieux rendre compte des avantages de la méthode en termes de qualité d'échantillon collecté, par rapport à un RDS simple.

## **Bibliographie**

Avin, C. and B. Krishnamachari. 2008, "The Power of Choice in Random Walks: An Empirical Study. *Computer Networks* 52(1), p. 44–60.

Beauchemin C. et al. *Trajectoires et Origines* (2016), « Enquête sur la diversité des populations en France", Ined Editions.

Marpsat M, Razafindratsima N. (2010), « Les méthodes d'enquêtes auprès des populations difficiles à joindre: Introduction au numéro spécial », *Methodological Innovations Online*, 5 (2), p. 3-16.

Mouw T, Verdery AM. (2012), «Network Sampling with Memory: A proposal for more efficient sampling from social networks». *Sociological Methodology*, 42(1), p. 206-256.

Schiltz M-A. (2005), « Faire et défaire des groupes : L'information chiffrée sur les «populations difficiles à atteindre» ». *Bulletin de Méthodologie Sociologique*, 86(1) p. 30-54.