

Table des matières

I-	Contexte général du projet de recherche ChIPRe.....	3
1-	Projet scientifique	3
2-	Contexte de l'immigration chinoise en France.....	3
a-	Une immigration ancienne	3
b-	Une population difficile à estimer	4
c-	Tendances récentes.....	4
3-	Composition de l'équipe et répartition des rôles.....	4
a-	Équipe de recherche.....	4
b-	Équipe du Service des enquêtes.....	5
c-	Équipe des enquêteur·rices.....	5
4-	Calendrier	6
II-	Présentation de la méthode Network Sampling with memory	7
1-	Une méthode de sondage par réseau	7
2-	Le fonctionnement de l'algorithme NSM	8
III-	La conception de l'enquête : de ChirDU à ChIPRe	10
1-	Champ de l'enquête ChIPRe.....	11
2-	Construction du Protocole	11
a-	La « Formative research »	11
b-	Enseignements du test de protocole en octobre 2019	13
c-	Mode de passation	16
d-	Protocole « pas-à-pas » à réaliser pour chaque enquêté·e	16
e-	Consignes pour les changements d'état	17
f-	Récapitulatif du circuit de collecte et des évolutions des états	18
3-	Conception du questionnaire et du roster	19
a-	Enseignements des tests	19
b-	Le questionnaire	19
c-	Le roster.....	21
4-	Mise en place et programmation des outils de collecte	22
a-	L'algorithme d'échantillonnage.....	22
b-	L'algorithme de matching.....	22
c-	Conception d'une application de collecte web	24
5-	Démarches CNIL	26
IV-	La préparation de la collecte.....	26

1-	Recrutement et formation des enquêteur·rices	26
2-	Les objectifs de collecte	26
3-	Organisation du terrain	27
4-	Outils de communication	27
a-	Objectifs.....	27
b-	Supports développés.....	27
c-	Moyens mobilisés en cours de collecte.....	28
V-	Bilan quantitatif de la collecte des données	28
1-	Déroulement du terrain	28
a-	Calendrier de la collecte et des confinements	28
b-	Calendrier des tirages.....	30
c-	Rythme de collecte	31
2-	Bilan en fin de terrain.....	33
a-	Chiffres clés	33
b-	État final du réseau (au 30 juin 2021)	34
c-	Profils des échantillonnés par rapport au réseau dévoilé.....	34
d-	Profils des échantillonnés selon l'issue	36
VI-	Les difficultés du terrain (bilan qualitatif).....	39
1-	Les implications d'une enquête par réseau	39
a-	La difficile transmission des réseaux entre enquêteur·rices.....	39
b-	La négociation des rosters.....	40
c-	Des réseaux précaires qui nécessitent une implication importante.....	42
d-	La gestion des graines complémentaires	43
2-	Des difficultés spécifiques à la méthode NSM	44
a-	Un algorithme qui nécessite de nombreux ajustements en cours de collecte	44
b-	Un algorithme d'échantillonnage qui produit des effets paradoxaux et désinsistatifs.....	45
c-	Et complexe à faire comprendre aux enquêté·es	46
3-	Les difficultés liées au contexte de la pandémie.....	47
a-	L'impact du COVID sur la collecte	47
b-	Le passage au distanciel	47
4-	Les évolutions des consignes au fil du terrain.....	48
a-	Les adaptations du protocole face aux réalités du terrain.....	48
b-	La gestion des rosters incomplets	48
c-	La gestion des refus.....	48
d-	La gestion du suivi dans l'application.....	48
e-	Le suivi de l'affectation des échantillonnés	49

5-	Les problèmes techniques.....	49
6-	Conclusion	49
VII-	Retours sur le questionnaire.....	50
1-	Les questions difficiles à faire comprendre.....	50
2-	Les questions sensibles	50
VIII-	Liste des annexes	51
	Annexe 1 : Réseaux des graines de l'enquête ChIPRe.....	51
	Annexe 2 : Schéma des résultats du test d'octobre 2019.....	51
	Annexe 3 : Délibération de la Cnil	51
	Annexe 4 : Plan des journées de formation	51
	Annexe 5 : Supports de communication	51
	Annexe 6 : Analyses statistiques menées sur la base des rosters.....	51
	Annexe 7 : Évolutions de la méthode d'échantillonnage.....	51
	Annexe 8 : Fonctionnement de l'algorithme de matching –présentation pour les enquêteur-rices	51

I- Contexte général du projet de recherche ChIPRe

1- Projet scientifique

Le projet « **Chinese Immigrants in Paris Region** », aussi appelé ChIPRe, a bénéficié d'un financement de l'Agence nationale de la recherche (ANR) et de l'Ined.

L'enquête ChIPRe est la première enquête quantitative en France spécifiquement consacrée aux immigrés chinois. Elle vise à :

- Tester pour la première fois en France la méthode Network Sampling with Memory (NSM) ;
- Évaluer la présence et la répartition des personnes nées en Chine vivant en Île-de-France ;
- Décrire leurs caractéristiques socio-économiques, leur parcours migratoire et professionnel, leurs conditions de vie en France et leurs relations avec la Chine ;
- Comprendre les éventuelles difficultés qu'ils rencontrent dans leur vie quotidienne, que ce soit concernant l'obtention d'un titre de séjour, leurs conditions de logement, la recherche d'un emploi, ou concernant les discriminations dont ils peuvent être victimes ;
- Décrire les réseaux de relations des personnes enquêtées selon leur région d'origine en Chine, pour étudier l'hypothèse qu'il existe une forte segmentation liée au parcours pré-migratoire des individus, à leur région d'origine en Chine et aux ressources et réseaux d'entraide mobilisables après l'installation dans le pays d'accueil dans le cadre d'une éventuelle communauté d'origine.

2- Contexte de l'immigration chinoise en France

a- Une immigration ancienne

La plupart des immigrés chinois sont arrivés en France à partir des années 1980. Il s'agit en partie de migrants économiques, originaires notamment du sud-est de la Chine (en particulier de la région de

Wenzhou ou de Qingtian au Zhejiang, ou du Guangdong), du Dongbei (nord-est) et des grandes métropoles chinoises.

L'immigration chinoise est désormais une composante importante de l'immigration en France. En 2017, la Chine se plaçait en 7^e position dans le classement des pays de naissance des immigrés non-européens et en 4^e position dans celui des admissions au séjour, derrière l'Algérie, le Maroc et la Tunisie.

b- Une population difficile à estimer

Près de 100 000 personnes nées en Chine ont été recensées en France métropolitaine en 2015. Mais des estimations indépendantes font état de chiffres beaucoup plus élevés : les immigrants originaires de la seule région de Wenzhou seraient plus de 110 000 à Paris (Li 2017) et autour de 250 000 en France (Béraha 2014). Par ailleurs, environ 50 000 Chinois et Chinoises seraient en situation administrative illégale (Lucchini 2012).

c- Tendances récentes

La population née en Chine enregistrée dans les recensements français a augmenté de près de 4% par an entre 2008 et 2015.

Des études qualitatives montrent une diversification des provinces d'origine en Chine.

Une part significative des immigrants chinois sont des étudiants en mobilité internationale : en 2015, les étudiants représentaient les deux tiers de l'ensemble des admissions au séjour de ressortissants chinois (contre 28% chez les autres ressortissants étrangers, toutes nationalités confondues). L'apport démographique des étudiants n'est toutefois que passager soit parce qu'ils quittent la France à l'issue de leurs études, soit parce qu'ils y demeurent sous un statut autre que celui d'étudiant. L'apport démographique des immigrants actifs est par contre a priori plus durable que celui des étudiants.

On assiste également à une féminisation croissante des flux. Les femmes représentaient plus de 55% des immigrants chinois recensés en France en 2015.

3- Composition de l'équipe et répartition des rôles

a- Équipe de recherche

En France :

- Isabelle ATTANÉ, directrice de recherche, responsable scientifique du projet ChIPRe
- Su WANG, chargée d'étude, post-doctorante : en charge de la « *formative research* », conception et traduction du questionnaire, référente de plusieurs enquêteur-rices sur le terrain, chargée de réaliser des questionnaires et des entretiens qualitatifs post-enquête
- Ya-Han CHUANG, chargée d'étude, post-doctorante : en charge de la *formative research*, conception et traduction du questionnaire, référente de plusieurs enquêteur-rices sur le terrain, chargée de réaliser des questionnaires et des entretiens qualitatifs post-enquête.

Aux Etats-Unis :

- Giovanna MERLI, démographe, Duke Population Research Center de l'Université Duke, co-responsable du projet, responsable de l'enquête ChIRDU (expérimentation NSM dans la région de Duke en 2018-2019 de laquelle s'inspire l'enquête ChIPRe)

- Ted MOUW, statisticien, Université de Caroline du Nord, co-concepteur de la méthode NSM
- Ashton VERDERY, statisticien, Pennsylvania State University, co-concepteur de la méthode NSM

b- Équipe du Service des enquêtes

- Gwennaëlle BRILHAULT, cheffe du service des enquêtes
- Aurélie SANTOS, ingénieure méthode : référente de l'enquête au SES, conception protocole avec le reste de l'équipe SES, suivi terrain
- Géraldine CHARRANCE, ingénieure statisticienne : traduction et programmation de l'algorithme d'échantillonnage en SAS, échantillonnages, pondérations, bilan et documentation
- Marine BONDON, ingénieure statisticienne : échantillonnages
- Paul COCHET, ingénieur statisticien : échantillonnages
- Kamel NAIT ABDELLAH, ingénieur informaticien : conception et programmation de l'application de collecte, conception et programmation de l'algorithme de matching
- Valérie LAPREE, assistante du SES : contrats des enquêteurs, suivi des états des sommes dues

c- Équipe des enquêteur·rices

<i>Équipe initiale - Septembre 2020</i>	<i>Équipe « bis » - Février 2021</i>
Shuyi LIU	Shuyi LIU
Yun LIU	Yun LIU
Yewei HU	départ
Ningsi SONG	départ
	Guangqi JIA
	Zhipeng LI
	Qian WAN
	Mingchao ZHOU

Tout au long de la collecte, les enquêteur·rices ont été supervisé·es sur le terrain par Ya-han Chuang et Su Wang, également en charge de réaliser des questionnaires. Ainsi, de septembre 2020 à février 2021 il y avait **2 référentes et 4 enquêteur·rices** sur le terrain, et de février à juin 2021, **2 référentes et 6 enquêteur·rices** dont 4 nouvelles·aux.

4- Calendrier

2019	1 ^e semestre	<p>Formative research (tout au long de l'année 2019) : repérage terrain, entretiens qualitatifs, identification des graines</p> <p>Visite à Duke University pour transfert de compétence sur NSM en février 2019</p> <p>Conception du questionnaire</p> <p>Traduction de l'algorithme NSM en SAS</p>
	2 ^e semestre	<p>Dépôt du dossier à la CNIL (juillet)</p> <p>Test en bureau du matching - aout 2019</p> <p>Test en bureau du dataflow entre application de collecte et échantillonnage SAS – septembre 2019</p> <p>Test questionnaire et du roster sur le terrain - octobre 2019</p> <p>Traduction du questionnaire en chinois</p>
2020	1 ^e semestre	<p>Informatisation et test en bureau des 2 versions du questionnaire (français et chinois)</p> <p>Recrutement des enquêteur·rices et préparation des formations</p> <p>Conception de l'application de collecte et de suivi en ligne Voozanoo 3</p> <p>Collecte prévue initialement pour mars → reportée en septembre en raison de la crise sanitaire</p>
	2 ^e semestre	<p>Début de la « pré-collecte » des graines – fin aout, début septembre 2020</p> <p>Formation des référentes le 2 septembre 2020</p> <p>Formations des enquêteur·rices les 14-15-16 septembre 2020</p> <p>Début de la collecte le 21 septembre 2020</p>
2021	1 ^e semestre	<p>Recrutement et formations d'enquêteurs en renfort, en cours de collecte. Formations BIS 8-9-10 février 2021.</p> <p>Fin de la collecte le 30 juin 2021</p>
	2 ^e semestre	<p>Apurement</p> <p>Pondération</p>

2022	1 ^e semestre	Bilan de collecte et documentation de l'enquête par le SES Valorisation méthodologique : présentation aux JMS Insee 2022 (SES) Premières analyses Giovanna Merli et Isabelle Attané, avec l'appui du SMS
	2 ^e semestre	Poursuites analyses : comparaison avec échantillon de chinois TeO2 (Isabelle Attané, Mathieu Ichou, SES) Valorisation méthodologique : participation au Symposium international de 2022 sur les questions de méthodologie organisé par Statistique Canada

II- Présentation de la méthode Network Sampling with memory

Après deux expérimentations menées à l'étranger en Tanzanie¹ et en Caroline du Nord (ChIRDU)² par l'équipe de recherche de Duke, conceptrice de la méthode et partenaire de l'enquête ChIPRe, la méthode a été utilisée pour la première fois en France entre septembre 2020 et juin 2021 dans le cadre de l'enquête ChIPRe. Cette partie vise à expliquer le contexte méthodologique dans lequel elle s'inscrit - à savoir les méthodes de sondage par réseau - et à expliciter les apports théoriques et le fonctionnement de l'échantillonnage NSM.

1- Une méthode de sondage par réseau

Les méthodes de sondage par chainage/réseau 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 principalement par trois raisons : (1) une croyance généralement acceptée selon laquelle les répondants sont plus susceptibles de participer et de répondre honnêtement à l'enquête lorsqu'ils sont désignés et invités à répondre par des connaissances ou des amis plutôt que contactés au hasard ; (2) un désir de comprendre la structure des réseaux et le rôle des relations sociales ; (3) une volonté d'enquêter, d'étudier et de comprendre des populations stigmatisées ou cachées, difficiles à atteindre avec les méthodes de sondage traditionnelles.

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 sous-jacentes sont les suivantes : (1) la population est formée d'un seul réseau : tous les individus sont connectés entre eux, plus ou moins directement. Cette hypothèse est essentielle car elle permet 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 : si A cite B, B citera nécessairement A. (3) Les individus recrutent aléatoirement parmi leurs connaissances.

¹ Merli M G, Verdery A., Mouw T, Li J., «Sampling Migrants from their Social Networks: The demography and Social Organization of Chinese Migrants in Dar es Salaam, Tanzania», Migr Stud. 2016 Jul;4(2):182-214.

² Merli G, Mouw T, Stolte A, Le Barbenchon C, Florey-Eischen F, Using Multiple Modes of Data Collection to Recruit Migrant Samples With Network Sampling With Memory: The Chinese Immigrants in Raleigh-Durham (ChIRDU) Study, 2019 Annual Meeting, Austin, TX April 10-13

Malgré cela, la précision de la méthode reste très sensible aux hypothèses qu'elle formule sur le réseau étudié (Gile et Handcock 2010³ ; Goel et Salganik 2010⁴). Un problème fondamental des approches fondées sur la marche aléatoire en général, et du RDS en particulier, est qu'elles ont une variance d'échantillonnage plus élevée que l'échantillonnage aléatoire simple. 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. Dans les réseaux présentant des degrés élevés de regroupement (clusters), le RDS a tendance à rester bloqué dans des composantes ségréguées du réseau, ce qui peut entraîner des échantillons biaisés. Dans certains réseaux, la variance d'un RDS pourrait être trop élevée pour être acceptable, mais le chercheur, ne connaissant pas le réseau dans lequel il échantillonne, serait incapable de distinguer les résultats corrects des résultats incorrects. Compte tenu de cette incertitude fondamentale, la capacité à déduire la prévalence de phénomènes mesurée par l'enquête au niveau de la population à partir d'échantillons fondés sur la marche aléatoire est discutable. En effet, Goel et Salganik (2010) ont trouvé des niveaux très élevés de variance d'échantillonnage pour le RDS dans un échantillonnage simulé utilisant des données de réseaux réelles.

2- Le fonctionnement de l'algorithme NSM

La méthode NSM a été créée par Ted Mouw et Aston Verdery et présentée dans un article de référence sur lequel nous nous appuyons ici pour décrire les principes de l'algorithme : Mouw and Verdery. 2012. "Network sampling with memory: a proposal for more efficient sampling from social networks." *Sociological methodology* 42(1).

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 dans le cadre de l'enquête, qui sont utilisées pour révéler progressivement la liste des membres de la population. Au fur et à mesure que l'échantillon progresse, la liste des personnes qui ont été 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 processus de sondage aléatoire simple.

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 surtout 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 : **un mode « List »** qui garantit des estimations asymptotiques sans biais (*Naive list mode* et *Even sampling*) et **un mode « Search »** qui est conçu pour fonctionner en tandem avec le mode « List » afin d'améliorer la précision des estimations.

L'intuition de base est assez simple : le processus d'échantillonnage est utilisé pour découvrir la liste des membres de la population, puis l'échantillonnage se fait dans cette liste avec remise en cherchant à favoriser des parties du graphe non encore explorées.

Chaque étape du processus d'échantillonnage consiste en un questionnaire avec un répondant qui, en plus des questions de fond de l'enquête, se voit demander de fournir les coordonnées de ses amis ou contacts qui sont membres de la population A (cette liste d'individus est 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 qui ont été désignés par les répondants.

³ Gile, K. J., & Handcock, M. S. (2010). Respondent-driven sampling: An assessment of current methodology. *Sociological Methodology*, 40, 285-327.

⁴ Goel S, Salganik M.J, «Assessing respondent-driven sampling», *PNAS* April 13, 2010 107 (15) 6743-6747

Les différents modes d'échantillonnage interviennent de manière successive au fil de la collecte : d'abord le *Naive list mode*, puis le *Search*, et enfin l'*Even sampling*. Afin de déterminer le mode d'échantillonnage à mobiliser, on s'appuie sur une batterie d'indicateurs permettant de rendre compte du niveau d'exploration du réseau. Parmi eux, on trouve la taille du réseau dévoilé (L), le nombre d'interviews réalisées (*Step*) et la part de personnes citées une seule fois parmi toutes les personnes dévoilées (P1).

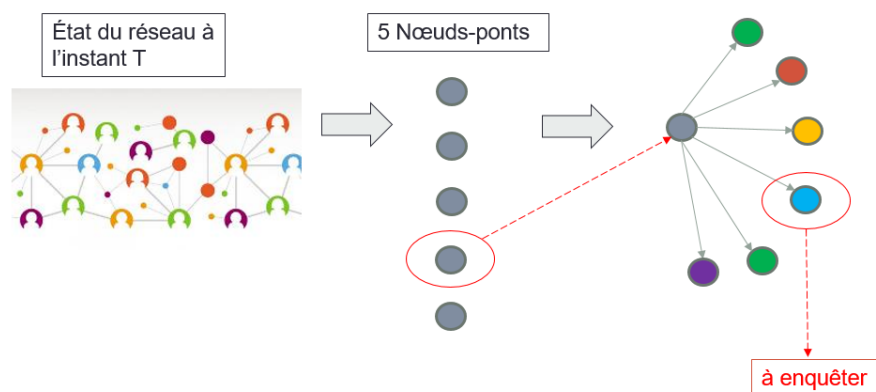
➤ ***Naive list mode : Démarrer la collecte aléatoirement***

La collecte démarre par l'identification de graines, qui seront les premiers individus enquêtés. Leurs réponses fournissent 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. Attention : le *Naive List mode* conduit à ce que les nœuds nommés en début de collecte vont être sur-échantillonnés par rapport à ceux apparus plus tard, les premiers apparus étant éligibles au tirage pendant plus longtemps.

➤ ***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 là d'un tirage à deux degrés. On mobilise ce mode après quelques interviews afin d'orienter au mieux l'échantillonnage vers de nouveaux pans du réseau, et éviter de rester bloqué dans un cluster.

Schéma simplifié du tirage en mode Search



➤ ***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 au cours du tirage. 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).

Les différents modes d'échantillonnage NSM et les critères de passage d'un mode à l'autre

Step	Mode			
	If Step<4			
1	Naive list			
2				
3				
	[Default mode]			
4	Search			
5				
....				
49				
50				
	If Step>50 & P1>A1 (=0.4)	If Step>50 & P1<A1 (=0.4)		
		L[Step-5]=L[Step]	L>200	L<=200 & (L[Step-5]<L[Step])
51	Search	Even Sampling	Even Sampling	Naive List
52				
...				

La sortie du mode *Search* repose sur deux critères :

- (1) avoir collecté plus de 50 questionnaires (Step>50)
- (2) lorsque le paramètre P1 (nombre d'individus cités une seule fois et non enquêtés) passe en dessous du seuil A1 (seuil fixé par les concepteurs de la méthode).

En pratique, le premier critère a été assez rapidement atteint. C'est le second critère (P1<A1) qui nous est apparu comme le plus discriminant, et qui nous a contraint à rester en mode *Search* jusqu'à la fin de la collecte. Ce critère se justifie de la façon suivante : lorsque P1 est bas, cela signifie que peu d'individus dans le réseau n'ont été cités qu'une fois, autrement dit, 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.

⇒ **Pour un bilan de la mise en œuvre de cet algorithme en collecte réelle, voir partie VI.1 et .2**

III- La conception de l'enquête : de ChirDU à ChIPRe

En février 2019, la référente du projet ChIPRe au SES (Aurélien Santos) s'est rendue à l'Université de Duke en Caroline du Nord afin de rencontrer les différents acteurs principaux impliqués dans la conception et la mise en œuvre de l'enquête ChIRDU. Cette visite a permis de mieux cerner le fonctionnement du tirage présenté ci-dessus, et de comprendre en détail le protocole mis en place sur le terrain. De retour de ce voyage, face à la complexité et à la dimension « artisanale » du protocole mis en œuvre, nous avons rapidement décidé de le repenser pour l'adapter au contexte français avec pour objectif de l'automatiser et de le simplifier. La partie III présente les choix méthodologiques opérés au cours de la conception du protocole ChIPRe et les ajustements réalisés par rapport à l'enquête de référence ChIRDU.

1- Champ de l'enquête ChIPRe

Population cible : les immigrés chinois nés en Chine âgés de 18 ans ou plus et vivant dans un département d'IDF, hors étudiants boursiers du gouvernement chinois.

L'exclusion des boursiers du gouvernement chinois est justifiée par la courte durée de leur projet migratoire : ils-elles rentrent généralement en Chine dès la fin de leur bourse.

Nous avons débattu de la pertinence d'ajouter une borne d'âge supérieure. Afin de favoriser l'enquête de personnes « actives » sur le marché du travail, l'équipe de recherche souhaitait instaurer une borne de 65 ans maximum. Nous avons renoncé à ce critère qui rendait l'éligibilité au roster trop complexe (inciter les enquêté-es à ne citer que des personnes de moins de 65 ans ?) et aurait impliqué une déformation non souhaitable du réseau des enquêtés. De plus, mobiliser des informations délivrées par un proxy pour juger de l'appartenance d'un individu au champ de l'enquête ne nous semblait pas judicieux car trop imprécis. Pour les mêmes raisons, nous avons renoncé à introduire un critère d'une durée minimale de 3 mois de résidence sur le sol français.

2- Construction du Protocole

a- La « *Formative research* »

L'exploration de terrain (« *formative research* ») est une composante essentielle de la méthodologie d'enquête *Network Sampling with Memory*.

Objectifs

Les objectifs de cette phase préliminaire sont nombreux :

- Préciser les critères qui serviront à stratifier l'échantillon des graines. D'après les enseignements de l'enquête ChIRDU, les critères principaux sont : âge, sexe, diplôme, durée de résidence dans le pays, région de provenance mais aussi « productivité de la graine » (peut-elle nous assurer de nous donner accès à un réseau suffisant ?). Voir si ces critères sont valables dans le contexte parisien et s'il faut en ajouter d'autres ? ;
- Trouver de futures graines. L'objectif est de sélectionner des graines dont les profils démographiques, les caractéristiques socioéconomiques, les parcours migratoires et les origines géographiques en Chine permettent de refléter au maximum la diversité de la population cible. (Cf. liste des types de réseaux ouverts par les graines en annexe 1) ;
- Trouver les lieux/personnes ressources : leaders associatifs, religieux, etc. ;
- Trouver les problématiques qui touchent cette communauté, qui les intéressent concrètement, afin d'adapter notre argumentaire et notre phase contact ;
- Calibrer les incitations financières ;
- Identifier des moyens de favoriser la coopération du répondant ;
- Définir quel moyen de communication est majoritairement utilisé dans cette communauté : téléphone, WeChat, mail ?

Méthode : Analyse du recensement + entretiens qualitatifs + visites de terrain

Des entretiens approfondis avec des individus appartenant à la population ciblée, afin de repérer les graines susceptibles de servir de point de départ à l'enquête, en ouvrant l'accès à une partie de leur réseau de sociabilité. Pour choisir nos graines, des échanges avec Giovanna Merli ont eu lieu à plusieurs reprises durant la phase de *formative research* afin de bénéficier de son expérience issue de l'enquête ChIRDU réalisée en Caroline du Nord.

Ces entretiens ont également eu pour but de tester l'argumentaire développé afin de convaincre les personnes sélectionnées de participer à l'enquête et d'ouvrir à leur tour l'accès à une partie de leur réseau de sociabilité.

Une vingtaine d'entretiens ont été menés sur le terrain par Su Wang et Ya-han Chuang, qui ont pu bénéficier des réseaux construits pendant leur travail de terrain de thèse. C'est un point important : la formative research est une étape encore plus lourde si l'équipe de recherche n'a pas d'ancrage préalable dans un réseau. Si nous avions dû partir de rien, cette étape aurait duré beaucoup plus de temps.

En plus des entretiens, une identification et une exploration des quartiers à forte concentration de population chinoise ont été réalisées. Isabelle Attané a observé, à partir des données des recensements de la population française, leur répartition géographique et leur niveau de concentration par Iris. Les quartiers/communes suivantes ont donc été explorés, et les visites ont fait l'objet de compte rendus de terrain : Crimée et Amériques (19eme), Ivry, Gennevilliers, Bagnolet, Créteil, Bussy Saint Georges, Livry Gargan, Bobigny, Aubervilliers.

Résultats

La phase exploratoire s'est révélée très riches d'enseignements. De multiples points de vigilance pour la collecte réelle ont été identifiés, et les échanges avec la population cible a nous ont apporté une meilleure compréhension des motivations des enquêtés.

Localisation :

- Il y a une diffusion importante en dehors de Paris intramuros ;
- Les arrivées les plus récentes sont dans le 91, le 92 et le 78 ;
- Il y a des immeubles quasiment exclusivement occupés par des chinois dans le 19^e ou à Gennevilliers par exemple. On peut constater parfois une forte suroccupation des logements avec le nombre de noms indiqués sur les boîtes-aux-lettres ;
- Mais difficile de savoir si le lieu de résidence est structurant dans le réseau (les enquêtés citent-ils leurs connaissances de quartier ?). Difficile de définir si c'est une caractéristique à intégrer dans le choix des graines ;
- Certains quartiers sont vides, occupés par des travailleurs aux horaires décalés, il faudrait y aller tard le soir pour les voir revenir du travail.

Personnes ressources / structuration du réseau :

- Il existe des agents immobiliers spécialisés dans la vente/location de biens pour les chinois ;
- Il existe des réseaux liés à l'église, peut-être plus d'orthodoxes que de bouddhistes ? L'entraide à l'arrivée peut passer par les réseaux de l'église ;
 - ⇒ Il peut être intéressant de trouver une graine dans ces réseaux.

Graines :

- Difficulté à trouver des graines très jeunes et des arrivées récentes, des hommes, et des sans papier ;
- Trouver des graines est extrêmement chronophage. Mais le temps joue aussi contre nous : plus on éloigne la collecte réelle de la rencontre des graines, plus il est difficile de maintenir le lien avec elles et de s'assurer qu'elles seront bien au rendez-vous au moment du lancement de la collecte ! De plus, la collecte a été décalée de mars 2020 à septembre 2020 en raison du confinement → Il a donc fallu redoubler d'effort pour ne pas perdre nos graines.

Motivation à participer :

- Au niveau des graines, pour les plus précaires, il faut beaucoup s'investir pour créer un lien, leur intérêt se situe dans le fait qu'ils voient Su et Ya-han comme des « ressources » pour les aider (préfecture, traduction, etc.). Nous avons donc des inquiétudes sur la manière de diffuser la confiance au-delà de l'échelon de la graine car ce type d'engagement ne sera pas possible pour les rangs plus avancés dans le réseau, il faut trouver des ressources.
- Giovanna Merli nous a apporté son expérience de ChIRDU : « il faut se demander « *What do they value ?* », chez nous c'était principalement l'éducation des enfants, et nous avons donc donné une adresse vers un service de conseil d'orientation pour l'éducation à Duke » → Nous n'avons pas les ressources pour créer de réel partenariat avec toutes les associations potentielles, nous créons à minima une plaquette qui rassemble les lieux de permanences juridiques et médicales avec un accueil en chinois à Paris.

Lieux de passation :

- Population très active, difficile de trouver du temps et un lieu adéquat pour faire les entretiens (donc les questionnaires)
- Très difficile d'enquêter les sans papier chez eux, trouver des lieux neutres ?

Information pour la collecte du roster et le matching :

- Problème : Certains ne connaissent pas le nom de leurs collègues car ils utilisent des pseudos « de travail ». La même personne peut donc être citée par son prénom d'une part, et par son pseudo d'autre part, le fait qu'il s'agit de la même personne peut échapper au matching.

Difficultés spécifiques pour l'accès aux sans-papier :

- Un enquête de *formative research*, sans-papier, nous renseigne : les sans-papiers ont peur qu'on rentre chez eux. De plus, le locataire ne veut pas qu'on découvre qu'il sous-loue son appartement. L'incitation financière ne peut pas régler ce problème, car ils ont très peur d'être découverts. Ils sont très vigilants et il ne suffit pas qu'ils soient cités par quelqu'un de leur réseau, ils peuvent soupçonner un piège, très difficile de faire confiance.
- Si on dit qu'on vient d'une institution de recherche nationale avec le papier officiel ? L'enquête signale que ça peut leur faire peur car ils ne lisent pas vraiment le Français et n'ont aucun moyen de vérifier en croisant les sources (par exemple vérifier le site de l'Ined etc.). Ils peuvent aussi se méfier s'ils pensent que ça vient de l'État.

b- Enseignements du test de protocole en octobre 2019

Stratégie pour les tests

Comment tester le protocole de manière optimale ? Réaliser un « pilote » (test d'une enquête miniature) sur le terrain implique nécessairement de « gâcher » des graines et leur réseau. De plus, ce pilote mené sur peu de cas ne permettrait pas forcément de tester les nombreuses dimensions du protocole et les multiples outils mobilisés. Nous avons donc privilégié :

1. Un test en bureau de toutes les dimensions qui peuvent l'être (appli de suivi, questionnaire informatisé, matching, échantillonnage, etc.)
2. Un test sur le terrain du questionnaire et du processus de négociation des rosters. En privilégiant des graines qui ne vont pas trop empiéter sur le réseau ciblé dans l'enquête réelle : personnes âgées, personnes nées à Taiwan ?
3. Un démarrage lent de l'enquête permettant d'ultimes ajustements, et permettant le lancement de quelques graines avant le démarrage du terrain avec les enquêteurs

Calendrier proposé :

- Aout 2019 : Des tests en bureau du matching avec des « faux rosters » pour caler les bonnes pondérations et les bons seuils
- Septembre 2019 : Des tests en bureau pour l'articulation suivante : Matching → export de la base Roster → échantillonnage NSM → ré-import dans l'appli VZN → actualisation des listings de suivi enquêteurs
- Octobre 2019 : Test du questionnaire et du roster :
 - Déterminer les difficultés de négociations des contacts dans le roster. Expérimenter et valider la « chaine de citations » suivante : *enquêter le citant → accepte de mentionner des cités → faisabilité de trouver et d'enquêter le cité.*
 - Déterminer la « réelle » durée de passation afin de revoir éventuellement la longueur du questionnaire avant son implémentation informatique.
- Novembre 2019 : Des tests en bureau du questionnaire en deux langues dans l'application avec différents scénarios pour passer par tous les filtres → *Retard, décalé à début 2020*
- Février-Mars 2020 : Commencer doucement l'enquête avec Su Wang et Yahan Chuang seulement. Faire les ajustements de protocole nécessaire en fonction des difficultés rencontrées. Puis faire la formation des enquêteurs et les lancer sur le terrain. → *En raison du confinement, étape décalée à septembre 2020*

Le test sur le terrain - Octobre 2019

Su Wang et Ya-han Chuang ont réalisé des tests sur papier en chinois auprès de 14 enquêté-es en suivant un réseau développé à partir de 3 graines. Les graines ciblées ont intentionnellement été des personnes nées à Taiwan ou Hong Kong et non en Chine, afin de préserver la population cible de l'enquête réelle. Bien que réalisé à partir de 3 graines et auprès de 14 enquêtés uniquement, le test nous a permis d'identifier de nombreux enjeux induits par cette méthode, et d'ajuster certains points du protocole en conséquence.

Les graines :

- Une graine sur trois n'a pas fonctionné. On risque d'avoir des « désistements » au moment du lancement de la collecte → Envisager d'en augmenter le nombre (attention, Giovanna Merli nous a recommandé de ne pas dépasser les 10 graines).
- Attention c'est aussi lié à la spécificité du choix de cette graine : on n'a pas pensé que l'influence de la politique est si profonde que les taiwanais refusent systématiquement les sujets portant sur la Chine ou les chinois.
- Plusieurs personnes rencontrées lors des tests ont montré de l'intérêt pour l'enquête et il leur a été proposé de devenir « graine » pour l'enquête réelle (elles étaient dans le champ de l'enquête, né.es en Chine).

Motivation à participer / type de liens et d'argumentaires à développer :

- Une enquêtée trouve que les Chinois d'ici sont plus pragmatiques et manquent de sens de participation sociale par rapport aux Chinois en Chine. Elle pense que la plupart de jeunes gens ne connaissent pas l'intérêt de l'enquête et n'associent pas leur propre situation au statut général des Chinois en France, il est donc difficile de mobiliser sur les arguments de conscience collective pour la communauté.
- Méthode assez intrusive. On repose sur le réseau interpersonnel, donc au-delà du type d'argumentaire, beaucoup de personnalisation est nécessaire autour de chaque questionnaire. Difficile de dire au revoir, il faut discuter beaucoup avant, après...

Les citations dans le roster :

- Le choix des personnes citées n'est pas aléatoire (hypothèse de départ de la méthode) mais s'oriente vers ceux qui sont les plus susceptibles de répondre. Biais. → Consigne enquêteur : expliquer à l'enquêté-e que l'on ne cherche pas des « candidats idéaux » à l'enquête, mais à connaître son réseau proche réel (Difficile à faire passer car les enquêté-es n'ont pas envie de déranger ceux qu'ils vont juger les plus « hostiles » au projet).
- Pour remplacer certains amis cités mais récalcitrants, des enquêté-es proposent d'ajouter a posteriori d'autres cités (Q11, Q13, Q14, Q15) → Nous avons pris conscience de cette possibilité grâce au test, mais nous avons conclu qu'il ne serait pas possible de laisser cette opportunité ouverte lors de la collecte réelle car elle conduit à une gestion de collecte trop complexe et à un risque d'erreur important lorsqu'un-e enquêteur-ice cherche à modifier les caractéristiques d'un cité pour saisir celles du remplaçant. De plus, cela perturbe la logique et les probabilités de tirage si le roster « initial » est déjà passé à l'échantillonnage et corrigé dans un second temps.
- On constate que les refus de citer des personnes dans le roster peuvent se présenter dès le 1^e rang (Q6 en R1), c'est-à-dire très proche de la graine, ce qui n'était pas réellement attendu.
- On constate que des citations doublons se présentent tôt, dès le rang 2.

Collecte des coordonnées WeChat/Tel et prises de contact :

- Obtenir le WeChat ID (permanent) et non le WeChat pseudo (temporaire) est un défi
- Une fois le WeChat obtenu, il faut que la personne accepte et nous « ajoute » dans son compte WeChat. Si elle refuse, on ne peut pas communiquer avec elle → Il faut rendre l'invitation convaincante !
- Pour maximiser ses chances de succès Ya-han a combiné les 2 approches : appeler d'abord puis envoyer l'invitation WeChat → Couteux : implique d'avoir collecté Tel+Wechat, et prend du temps
- Pour augmenter ses chances de succès (que le cité accepte l'invitation WeChat), il faut l'aide du citant en parallèle qui prévient ses amis. Il faut donc leur expliquer comment communiquer sur l'enquête auprès de leurs amis. Prends du temps !
- Les personnes âgées (du test) sont ancrées depuis plus longtemps en France et ont des réseaux solides, et ils ont aussi plus de temps, il est donc plus facile d'obtenir des citations de rosters de leur part. De plus ils fonctionnent tel et non WeChat, ils sont donc plus faciles à joindre directement (pas d'étape d'acceptation de l'invitation WeChat) → Attention : la collecte des rosters et les prises de contact seront potentiellement plus ardues avec d'autres profils.

La gestion de notre(nos) compte(s) WeChat :

- Une enquêtée nous a alerté sur le côté « suspect » de la communication autour de l'enquête, elle a cru que c'était une arnaque sur WeChat (fréquentes par ailleurs). Notre message et image ne font pas très officiels. Tout est très serré sur le message. Le message sur les bons d'achat fait penser que c'est un piège. Le nom du groupe n'inspire pas confiance "groupe d'enquête numéro 1". Enfin, attention, WeChat a inversé des phrases donc le message avait peu de sens !
- Une autre enquêtée : pas de cercles d'amis dans notre WeChat, ce qui le rend peu crédible
→ Il faut alimenter le compte WeChat avec des publications régulières
→ Il faut mettre une photo de profil des enquêteurs
→ Il faut mettre le logo de l'INED quelque part (photo du compte ou sur l'annonce) et mettre dans une seule image le logo, le texte, le lien vers le site internet

Les bons cadeaux :

- Établir un protocole clair sur le nombre minimal de coordonnées WeChat ou tel à recueillir pour recevoir le bon cadeau
- Pas toujours évident de faire accepter les tickets cadeau, certains magasins refusent → Mettre à disposition sur le site la liste des magasins où utiliser les tickets, les refus en magasins risquent de peser sur la réputation de l'enquête
- Réception des tickets : Pour une dame à la retraite, le ticket a été un argument qui a pesé dans la négociation. Pour les employés ou étudiants : cela n'est pas assez. Il y a aussi eu 1 personne méfiante qui trouve étrange de donner de l'argent. Généralement, ça ne sert pas forcément à convertir des gens à accepter le roster mais c'est « sympa » → Diversité de réaction, mais utile pour les plus précaires, on maintient les incitations financières

Autres infos techniques :

- Nécessité d'importer la langue chinoise dans tous les téléphones avant le début de l'enquête réelle.
- Attention il faut que les téléphones pour l'enquête réelle puissent lire les QR code pour importer des fiches de contact WeChat le cas échéant
- Que faire des « 99 » dès les variables de nom ou de prénom pour le matching ?

⇒ *Pour plus d'informations, voir le schéma du test en Annexe 2*

c- Mode de passation

Questionnaire informatisé disponible en français ou en chinois sur une application de collecte web dédiée, nécessitant une connexion à internet via les smartphones (Samsung A10) fournis par l'Ined.

Passation des questionnaires en face-à-face de préférence. Puis passation par téléphone ou en visio compte tenu du contexte sanitaire au moment de la collecte.

d- Protocole « pas-à-pas » à réaliser pour chaque enquêté-e

	Face-à-face	Distanciel
1	Listing « à enquêter » : consulter la liste des personnes à enquêter, faire les relances, noter les évènements de contact, poser les questions d'éligibilité, fixer un RDV...	
2	Lors du RDV pour la réalisation du questionnaire : vérifier bien que vous êtes en présence de la bonne personne « à enquêter » (onglet <i>Ident et Consent</i> de l'appli). Saisir le bon RID au début du questionnaire.	
3	Lire à haute voix le consentement pour la passation du questionnaire et le compléter (date + RID)	Lire + compléter le consentement papier. Vous pouvez montrer le consentement en PDF partage d'écran pour que l'enquêté-e puisse suivre avec vous, en cas de visio
4	Remettre à l'enquêté-e une version papier du consentement pour le questionnaire. Et conserver un exemplaire.	Faire envoyer cet exemplaire par voie postale avec les chèques cadeaux. Conserver un exemplaire.
5	Réaliser la passation du questionnaire. Saisir les réponses dans l'appli. Montrer le cahier des cartes codes.	Réaliser la passation du questionnaire. Saisir les réponses dans l'appli. Envoyer les cartes codes en PDF avant la passation, ou en partage d'écran dès que nécessaire

6	Remettre le chèque-cadeau de 15€ pour la passation du questionnaire (même si celui-ci n'a pas été rempli jusqu'à la fin)	Faire un mail à Isabelle après la passation, pour l'envoi postal du/des chèques cadeaux
7	Négociation de la collecte du roster, et expliquer les critères d'éligibilité au roster	
8	Lire consentement au roster et le compléter (date + RID) puis le remettre à l'enquêté et conserver un exemplaire	Lire le consentement au roster et le compléter (date + RID) + montrer en partage d'écran en cas de visio. Envoi postal. Et conserver un exemplaire
9	Négociation des coordonnées des amis cités	
10	Remettre le chèque cadeau de 20€ pour la passation du roster si a donné au <u>minimum 3 coordonnées de contact</u>	Faire un mail à Isabelle après la passation, pour l'envoi postal des chèques cadeaux
12	Après la passation : Mettre à jour les Listings sur l'application : <input type="checkbox"/> Listing Interviews pour la mise à jour de R_TICKET <input type="checkbox"/> Listing Incomplets pour la mise à jour des coordonnées de contact reçues post passation	
13	Valider le roster de l'enquêté une fois qu'il est complet en cochant Statut Matching = 1 pour toutes ses lignes → Le roster part à l'échantillonnage → vous recevez de nouveaux individus « à enquêter » dans votre listing (le mardi suivant)	

e- Consignes pour les changements d'état

Passage à l'état « Hors-Champ » :

- ✓ Moins de 18 ans
- ✓ Pas né-e en Chine
- ✓ Étudiant-e boursier du gouvernement chinois
- ✓ Réside hors de l'Île-de-France
- ✓ A quitté l'Île-de-France depuis qu'il-elle a été cité-e

Passage à l'état « Impossible à enquêter » :

- ✓ Personne qui ne peut communiquer avec l'enquêteur-riche du fait de la barrière de la langue (dialecte)
- ✓ Ou qui est en incapacité physique ou mentale de mener l'entretien.
- ✓ Ou si vous ne vous sentez pas en sécurité à l'idée de vous rendre seul-e chez cette personne. Et qu'elle refuse une rencontre en dehors de son domicile

Passage en « refus » :

- Consigne initiale

Quelqu'un qui refuse de participer à l'enquête au bout de : 3 relances à 3 jours d'intervalle + pause de quelques semaines + 1 relance.

Toutefois cette consigne a été établie arbitrairement, sans visibilité sur le rythme de collecte induit par la méthode NSM, et sans recul sur les taux de refus des personnes citées par leurs pairs. Peu après le début du terrain, dès le mois d'octobre, nous avons dû ajuster certaines consignes.

➤ Consignes en cours de terrain

Les refus explicites : vous devez estimer si ce refus est vraiment définitif pour le classer immédiatement en refus. Sinon vous pouvez le conserver et faire 1 tentative de relance pour voir si la personne, malgré un 1^e refus exprimé, a pu changer d'avis (être mieux disposé, dans une autre humeur, moins occupé) quelques semaines plus tard.

Les refus non explicites : esquive/évitement, pas de validation d'invitation WeChat, multiples lapins...
On estime que pour classer en refus il faut :

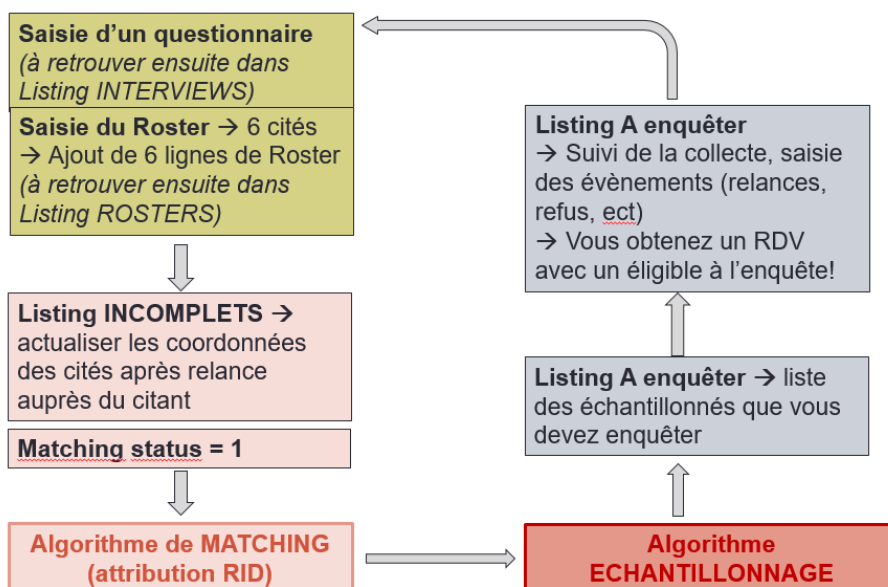
- 2 semaines d'exploitation avec plusieurs tentatives de contact et de négociation saisis dans l'application de suivi (3 relances à 3 jours d'intervalle est indiqué sur le PPT de la formation de février 2021, mais ce n'est pas aussi formalisé lorsque l'on évoque cette règle en octobre 2020). ATTENTION : 2 semaines = à partir du 1^e contact et non pas à partir de la date d'échantillonnage.
- À l'issue des 2 semaines, il faut trancher en concertation avec Isabelle (Aurélie en appui si besoin). Si toutes les personnes en attente basculent automatiquement en refus au bout de 2 semaines sans concertation, ça risque d'augmenter de beaucoup le taux de refus.

⇒ **Voir également partie VI.4 sur les évolutions de consignes au cours de la collecte**

f- Récapitulatif du circuit de collecte et des évolutions des états

Au fil de l'enquête, chaque individu cité (par un précédent enquêté) puis échantillonné devient à son tour un individu potentiellement enquêté. Ce cycle fait transiter chaque individu présent dans la base ROSTER par divers états qui varient au fil de la collecte selon le cycle suivant :

Récapitulatif des étapes du « cycle de collecte » (source : formation enquêteurs)



Récapitulatif des états qui peuvent être attribués à chaque ligne de la base roster (source : formation enquêteurs)

N°	Etat (rstatus)	Explication - récap
0	Incomplète	État initial par défaut d'une ligne de roster créée. Automatique.
1	Complète 😊	Étape de la relance au citant pour obtenir les coordonnées du cité: → soit vous les obtenez (état 1) → soit vous ne les obtenez pas (état 2)
2	Sans contact	Ces états se mettent à jour <u>automatiquement</u> quand vous cochez "matching status = 1.prête pour le matching" (c'est-à-dire quand vous estimez avoir TOUT fait pour obtenir les coordonnées de ce cité)
3	A enquêter (SC)	Étape de l'échantillonnage: Ces états sont mis à jour <u>automatiquement</u> . Les échantillonnées apparaitront directement dans votre listing "à enquêter". (état 3 = état 2 échantillonné)
4	A enquêter	
5	Hors-champ	Étape du suivi de collecte: Certains de vos "à enquêter" deviendront "hors-champ", "impossible" ou "refus" (consultez nous pour les refus).
6	Impossible	→ C'est à <u>VOUS</u> de mettre à jour ces états ⚠
7	Refus	
8	En cours	Étape d'enquête: certains heureusement seront enquêtés!
9	Enquêté 😊	Cet état est mis à jour <u>automatiquement</u> par l'application lorsque vous cliquez sur « enregistrer » dans l'onglet Fin du questionnaire

3- Conception du questionnaire et du roster

a- Enseignements des tests

Lors de la conception du questionnaire, Su Wang et Ya-han Chuang ont effectué quelques tests papier (mai 2019) pour un premier aperçu de l'acceptabilité des thématiques du questionnaire et du roster, et la cohérence interne du questionnaire. Puis lors du test sur le terrain en octobre 2019, nous avons identifié un certain nombre de modifications à apporter au questionnaire. Les parties EMPLOI (modules I et J), DISCRIMINATIONS (module L) et RELATIONS SOCIALES (module M) sont celles qui ont connu le plus d'ajustements post-tests. Nous avons également noté une vigilance à avoir sur certaines parties du questionnaire qui peuvent aboutir à des récits fleuves sur leur trajectoire migratoire de la part des enquêtés.

En février 2020 : les tests du questionnaire informatisé ont été réalisés en interne par l'équipe de recherche et le SES avec pour objectif de vérifier la cohérence des filtres, la fluidité de passation et la cohérence entre les versions française et chinoise du questionnaire.

b- Le questionnaire

Le questionnaire a principalement été construit par Isabelle Attané, Su Wang, Ya-han Chuang, avec l'appui d'Aurélien Santos au SES. Dans sa version finale, il s'éloigne beaucoup du questionnaire initial de l'enquête CHIRDU. La trajectoire d'emploi y est notamment beaucoup plus développée (dernier emploi avant de quitter la Chine, 1^{er} emploi en France, emploi actuel ou dernier emploi occupé). Les questions sur les relations avec les diverses communautés chinoises résidant en France sont également plus approfondies. En revanche, les questions sur la santé sont moins nombreuses.

Plan du questionnaire final

Le questionnaire est composé des 14 modules suivants :

- A. Caractéristiques personnelles** : âge, sexe, date et lieu de naissance, etc.
- B. Couple et famille** : Statut matrimonial, nombre d'enfants, etc.
- C. Parcours scolaire & Éducation** : Pays d'étude, diplômes obtenus, etc.
- D. Conditions de logement et cadre de vie** : Habitants et surface du logement, informations sur le quartier, etc.
- E. Situation pré-migratoire** : Résidence en Chine, pays de naissance des parents, situation professionnelle avant la migration, etc.
- F. Conditions d'arrivée et situation administrative** : Année d'arrivée en France, nationalité, etc.
- G. Parcours migratoire** : Raisons de la migration, organisation du voyage pour venir en France, etc.
- H. Usage et maîtrise du français** : capacité à lire, à parler, à écrire, langue parlée au domicile et sur le lieu de travail, etc.
- I. Premier emploi en France** : profession, type de contrat, secteur, etc.
- J. Emploi** : l'enquêté-e passe soit par J1 soit par J2:
- J1 : si en emploi au moment de l'enquête : les questions concernent l'emploi actuel (profession, type de contrat, secteur, activités complémentaires, revenus)
- J1_1: sous-module focus sur les employé-es
- J1_2: sous-module focus sur les chef-fes d'entreprise
- J2 : si pas en emploi au moment de l'enquête : les questions sont les mêmes que dans J1 mais concernent le dernier emploi occupé en date (avant la retraite ou avant la période de chômage)
- J2_1: sous-module focus sur les ancien·nes employé-es
- J2_2: sous-module focus sur les ancien·nes chef-fes d'entreprise
- K. Conditions de vie et statut social** : état de santé, conditions de travail, pouvoir d'achat, sentiment d'intégration, etc.
- L. Discriminations** : déclaration des agressions, racismes, discriminations, dans diverses sphères sociales, etc.
- M. Relations sociales** : description du réseau de 1er et 2nd cercle, relations avec des Chinois d'autres origines régionales, vie associative, etc.
- N. Liens avec la Chine** : relations avec des membres de la famille restés en Chine, envoi d'argent.

Les cartes-codes

Six questions étaient accompagnées de carte-codes à montrer à l'enquêté-e pour faciliter la passation :

- *g_migrais* : Les raisons de la migration (carte n°1)
- *g_migorg* : l'organisation du premier départ de la Chine vers la France (carte n°2)
- *j1_unempdiscri* : les raisons pour lesquelles l'enquêté ne trouve pas d'emploi (carte n°3)
- *j1_revfract* & *j2_revfrader*: les tranches de revenu mensuel (carte n°4)
- *j1_actypact* & *j2_actypder* : les activités complémentaires à l'emploi principal (carte n°5)

- *i_pqagress, i_pqracist & i_pqdiscri* : les motifs de discrimination (carte n°6)

c- Le roster

À l'issue du questionnaire, l'enquêté répond (s'il accepte) à une seconde partie appelée « roster », qui sert à collecter des informations sur 6 membres de son réseau de sociabilité.

Dans CHIRDU, il y avait 3 rosters (Chinois locaux / Non chinois locaux / Chinois hors US). Nous avons maintenu uniquement le 1^{er} roster concernant les relations entretenues par l'enquêté-e avec des personnes chinoises résidant en Ile-de-France.

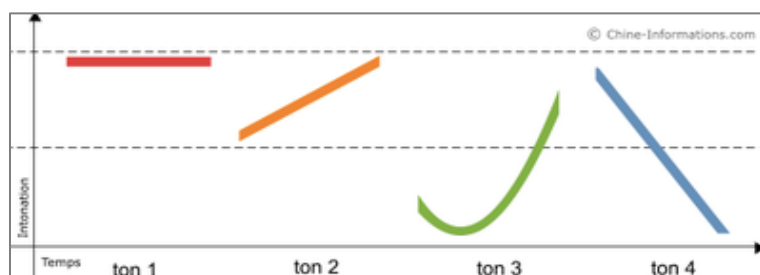
Pour chaque personne citée par l'enquêté-e, 12 questions sont posées afin de collecter des informations (identifiantes) permettant de joindre le cité (si échantillonné) et de décrire le lien/la relation entre le citant et le cité. Les questions sont les suivantes :

Nom de famille en pinyin + ton*
Initiales du prénom chinois + ton*
Pseudonyme français ou chinois (si la personne en utilise un)
Type de relation : Cette personne est ? ami proche, connaissance, famille, collègue, voisin, autre
Sexe : Cette personne est un homme ou une femme ?
Province d'origine en Chine (d'après ce que vous savez) (menu déroulant)
Âge : Dans quelle tranche d'âge se situe cette personne ? (10 tranches)
Fréquence des échanges : Parlez-vous à cette personne...: (5 modalités)
Durée de résidence en France (d'après ce que vous savez): (8 tranches)
ID WeChat (permanent)
Pseudo WeChat (temporaire)
Numéro de téléphone

* Pour garantir l'anonymat du cité, nous avons choisi d'utiliser le pinyin complet du nom de famille (peu de diversité de nom de famille en chinois, donc peu identifiant) auquel est ajouté le « ton » (cf. schéma suivant), et les initiales de chaque syllabe du prénom + le ton. Dans les cas où le prénom ne comporte qu'une voyelle, alors elle est également notée avec le ton associé.

- Ex. 徐封 → Xu2 F1
- Ex. 陈安 → Chen2 An1
- Ex. 张天一 → Zhang1 T1Y1

Schéma représentant les 4 intonations utilisées dans la langue chinoise



4- Mise en place et programmation des outils de collecte

a- L'algorithme d'échantillonnage

Programmation de l'algorithme : Géraldine Charrance, SES.

Suite au voyage à Duke en 2019, nous avons décidé d'adapter le programme fourni par le concepteur de la méthode pour le rendre opérationnel dans le cadre de l'enquête ChIPRe. Nous avons fait le choix (la statisticienne, avec l'appui de la référente du projet et de la cheffe du SES) de repartir de « zéro », c'est-à-dire de :

- 1) Comprendre les principes de l'algorithme en s'appuyant sur l'article théorique fondateur de la méthode⁵
- 2) Écrire un nouveau programme en SAS, qui traduirait ces principes théoriques.

Ces deux étapes ont constitué un lourd travail de plusieurs mois pour la statisticienne en charge de la programmation puis de l'utilisation de cet algorithme, avec des occasions très restreintes de dialoguer avec le concepteur initial de la méthode.

⇒ **Le fonctionnement de l'algorithme est décrit en partie II.2**

b- L'algorithme de matching

Programmation de l'algorithme : Kamel Nait Abdellah, SES.

Comme mentionné en introduction de cette partie III, nous avons cherché à automatiser un certain nombre d'opérations réalisées manuellement dans le protocole ChIRDU. Nous avons ajouté une étape standardisée de matching précédant le tirage.

Le matching permet d'identifier les individus cités une ou plusieurs fois dans le réseau en leur attribuant un identifiant appelé RID. Il constitue une aide à la décision dans l'identification de potentiels doublons dans la base des rosters.

⁵ Mouw and Verdery. 2012. "Network sampling with memory: a proposal for more efficient sampling from social networks." *Sociological methodology* 42(1).

<div style="border: 1px solid black; padding: 5px; width: fit-content; margin-bottom: 10px;"> Les 9 variables utilisées lors du matching </div>	R_NOM	Nom de famille en entier en pinyin + ton Ex : WANG1 pour Wang 1 ^{er} ton
	R_PREN	Initiales du prénom chinois en pinyin + ton Ex: X3T4 pour Xiatong 3 ^e et 4 ^e tons
	R_PSEUDO	Pseudo français ou chinois. Si l'enquêté utilise les deux dans la vie courante, noter le pseudo ou le prénom français.
	R_REL	Type de lien entre l'enquêté et le cité: Amie proche, Connaissance, Membre de la famille, Collègue, Voisin, Autres
	R_SEX	Sexe de la personne citée
	R_PROV	Province d'origine en Chine (menu déroulant)
	R_AGE	Tranche d'âge
	R_PARL	Fréquence des relations avec le cité: se parlent chaque jour, semaine, mois, plusieurs fois par an...
	R_DUREE	Nombre d'années de résidence en France (tranches de 5 ans)
	R_WC ⚠	Identifiant WeChat unique du cité (<u>par lequel il pourra être joignable pendant toute la durée de l'enquête</u>) <i>Ne pas confondre avec R_WCPSEUDO</i>
	R_WCPSEUDO ⚠	Pseudo WeChat actuel du cité
	R_TEL	Numéro de téléphone portable du cité

Le principe est le suivant :

- Les variables d'identification de chaque ligne de la base des rosters (voir ci-dessus : R_NOM, R_PREN, R_PSEUDO, R_SEX, R_PROV, R_AGE, R_DUREE, R_WC et R_TEL) sont comparées une à une à celles de chacune des lignes existantes et disposant d'un identifiant (RID affecté et validé). Des taux de **similarité variable par variable** sont ainsi calculés.
- Un taux de **similarité globale** est ensuite calculé pour chaque paire afin d'identifier la ligne la plus proche de la ligne en cours de matching. Selon les valeurs de ces taux de similarité, l'algorithme affecte soit un RID existant à la ligne courante, soit un nouveau RID.
 - o Les lignes ayant **un score élevé** se ressemblent et sont susceptibles d'être des doublons → l'algorithme attribue un RID identique = la même personne a été citée 2 fois dans le réseau
 - o Les lignes ayant **un score faible** sont différentes donc uniques → si toutes les paires sont dans ce cas, l'algorithme attribue un RID nouveau = une nouvelle personne, citée pour la 1^{ere} fois, apparait dans le réseau
 - o Pour les lignes obtenant **un score intermédiaire**, nous avons ajouté une étape de « vérification humaine » (s'agit-il de doublons ou non ?). Cette étape nécessite souvent un travail d'investigation de la part des enquêteurs pour déterminer la probabilité que ces lignes désignent ou non une même personne, d'après leur connaissance du réseau. L'étape de vérification était réalisée chaque semaine par Aurélie Santos, via le listing MATCHING dans l'appli, qui permettait de visualiser tous les cas ayant obtenu un score intermédiaire.

Calcul des taux de similarité entre les variables une à une

- La comparaison des variables textuelles (R_WC, R_NOM, R_PREN, R_PSEUDO, R_TEL) calcule un taux de similarité en se basant sur la mesure de la distance de Levenshtein.
- La comparaison des variables modales (R_AGE et R_DUREE) calcule un taux d'adjacence entre les valeurs modales. Ce taux est calculé en se basant sur la différence des valeurs comparées

de la variable modale et de l'étendue de cette variable. Plus la différence est petite, plus le taux de similarité sera grand. Lorsque les valeurs sont identiques, le taux est de 100 %.

- La comparaison des valeurs de la variable de province d'origine (R_PROV) est basée sur l'égalité entre les valeurs modales. Si celles-ci sont égales (donc même province), le taux de similarité sera de 100%, autrement, ça sera 0%.

Vu l'obligation de saisie pour chaque variable du roster utilisées dans le matching, celles-ci ne sont jamais vides. Cependant, en présence de réponses en 44, 88_REFUS ou en 99_NSP, le taux de similarité sera de 0% pour la variable concernée.

Calcul du taux de similarité globale entre les lignes

Fin 2019 : lors des tests en bureau, les rosters collectés sur le terrain lors du test en octobre ont été saisis dans l'application web de collecte afin de tester l'algorithme de matching.

Après ces simulations, nous avons retenu la formule suivante pour le calcul du score de similarité globale :

$$S = 0,30 * \text{MAX}(s(R_WC), s(R_TEL)) + 0.175 * (s(R_NOM) + s(R_PREN)) + 0,07 * (s(R_PSEUDO) + s(R_SEX) + s(R_AGE) + s(R_DUREE), s(R_PROV))$$

Nous avons choisi de donner davantage de poids dans le score aux variables particulières identifiantes (R_WC et R_TEL), en ne retenant que celle qui obtient le meilleur score car d'une part, elles ne sont que rarement toutes les deux renseignées, et d'autre part, si on obtient un score de similarité de 100% sur une des deux, alors, il n'y a quasiment aucun doute sur le fait que ce soit la même personne. On ne veut donc pas que le score baisse en raison d'une faute de frappe ou information non renseignée, car ces informations sont en théorie uniques et ne peuvent être identiques pour deux individus différents.

- ⇒ À la suite de cette étape de matching, réalisée avant chaque tirage, le roster général (dont chaque ligne porte un RID) peut être exporté de l'application en ligne et être soumis à l'algorithme d'échantillonnage SAS.

c- Conception d'une application de collecte web

Conception de l'application : Kamel Nait Abdellah pour la conception et la programmation. Aurélie Santos et Géraldine Charrance en appui pour la conception.

Objectifs de cet outil : Centraliser les informations et combiner de multiples fonctions :

- Centraliser la collecte des questionnaires
- Réaliser le matching
- Réaliser le suivi de collecte (saisie des événements de contact avec les échantillonnés)
- Suivre l'évolution du réseau en temps réel
- Veiller collectivement aux erreurs d'identification des doublons

Principaux listings utilisés dans le suivi de collecte

LISTING	FONCTION
Interview	<ul style="list-style-type: none"> - Savoir combien de personnes ont déjà été enquêtées - Avoir accès en un clic aux profils des enquêtés (sexe, province de naissance, etc)

	<ul style="list-style-type: none"> - Retrouver un questionnaire à l'état « 8. En cours » (s_etat) pour le compléter en cliquant sur « edit » (e) en début de ligne - Effectuer le suivi des tickets à transmettre (r_ticket) - Voir le nombre de lignes de rosters collectées par questionnaire (r_lignes)
Roster	<ul style="list-style-type: none"> - Voir toutes les personnes citées par tous les enquêté-es (provenant de tous les enquêteurs) : c'est le réseau que l'on explore et qui est complété au fil de la collecte - Les lignes faisant référence à des personnes citées plusieurs fois portent le même RID (après le matching)
Incomplet	Voir les lignes de roster incomplètes, pour lesquelles il manque les infos de contact (état « 0. Incomplet »). Il permet donc de retrouver facilement les enquêté-es que les enquêteur-trices doivent relancer pour obtenir les coordonnées des personnes qu'ils/elles ont citées
À enquêter	<ul style="list-style-type: none"> - Chaque enquêteur voit la liste des personnes échantillonnées qui lui ont été attribuées - Il peut effectuer le suivi de la collecte: noter les prises de contacts auprès de ces personnes ou auprès de leur citant - Il peut voir en un coup d'œil le dernier évènement saisi (REF = 1 indique la ligne dans laquelle effectuer le suivi en cas de doublon, lorsque deux lignes ou + portent le même RID) - Il doit mettre à jour les états suivants, s'il se retrouve dans ces situations : <ul style="list-style-type: none"> 5. Hors-champ (individu ne faisant pas partie du champ de l'enquête) 6. Impossible (impossible à enquêter) 7. Refus (refuse de répondre)

Interface de saisie d'évènement dans le listing « à enquêter »

Historique des événements de suivi et des états

ligne de référence pour le suivi

Oui

Évènements

1- bloc de saisie d'un nouvel évènement affiché seulement pour une ligne référente

Remarque : tout évènement saisi sera historisé après sauvegarde de cette fiche et ne sera plus modifiable

Date évènement

30 03 2020

Heure évènement

Description évènement

évènements historisés (suivi)

2. Saisie historisée après sauvegarde et les champs réinitialisés pour une prochaine saisie. Dernier évènement saisi affiché en haut.

Date évènement

Mars 2020

Lu	Ma	Me	Je	Ve	Sa	Di
24	25	26	27	28	29	1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31	1	2	3	4	5

États

états historisés

bloc toujours affiché présentant l'évolution des états d'une fiche au cours de l'enquête. Toutes les fiches partageant le même RID auront le même contenu de l'historique des états dupliqué à partir de la ligne référente.

5- Démarches CNIL

Cette enquête a fait l'objet d'une saisine DPD (Délégué à la Protection des Données), d'une Analyse d'Impact relative à la Protection des Données (AIPD) et d'un dossier soumis à la CNIL.

L'AIPD a été nécessaire car l'enquête comprenait deux dimensions spécifiques : une dimension innovante avec l'algorithme d'échantillonnage NSM, et elle portait sur une population cible vulnérable.

Suite au dépôt du dossier CNIL en juillet 2019, s'en est suivi de nombreux échanges avec la CNIL pendant 7 mois. La délibération finale de la Cnil, éditée le 27 février 2020 et reçue le 04 mars 2020 à l'Ined, est en annexe 3. Les points 11, 14, 19, 20, 21 constituent des réponses au débat lié aux rosters. Enfin, la délibération fait émerger une nouvelle problématique spécifique à l'enquête : le transfert de la base de données hors de l'UE pour l'exploitation des données par l'équipe partenaire résidant aux États-Unis (point 27). Ce type de transfert impose de compléter un formulaire annexe spécifique et de mettre en œuvre des mesures de sécurité supplémentaires comme des clauses contractuelles type (toujours en cours d'instruction au moment de la rédaction du bilan).

IV- La préparation de la collecte

1- Recrutement et formation des enquêteur·rices

Le recrutement s'est fait par la diffusion d'une fiche de poste dans divers réseaux, en en parlant aux enquêtrices sinophones recrutées en 2020 dans le cadre de l'enquête TeO2 pour voir si elles connaissaient des personnes intéressées, et par la recherche dans le réseau de connaissance de Su Wang et Ya-han Chuang.

Il s'agissait d'un recrutement complexe aux exigences multiples : enquêteurs familiarisés avec le terrain, sinophones, idéalement certains parlant le dialecte Wenzhou, bien intégrés dans la communauté.

La complexité du terrain, renforcée par le confinement, nous a conduit à étendre la période de collecte. Deux de nos enquêteur·rices initiales aux n'étant pas disponibles au-delà de février 2021, nous avons dû procéder à une seconde vague de recrutement et de formation, en cours de collecte. Au final, il y a donc eu 2 sessions de formation : en septembre 2020 et en février 2021.

Les formations se sont déroulées sur 3 jours (ou 2,5 jours pour la 2^e session, plus condensée) en raison de la quantité d'information à transmettre et de la complexité du protocole que les enquêteur·rices devaient s'approprier. Le premier jour, nous avons choisi de présenter d'abord l'enquête de manière générale et contextualisée, puis de présenter le questionnaire module par module. La seconde journée était consacrée à l'application de collecte et de suivi, puis aux exercices de mises en situation de passation en petits groupes. Le 3^e jour a permis d'expliquer l'organisation du travail et d'autres détails pratiques, et de faire un quiz récapitulatif des enseignements des deux précédentes journées.

⇒ *Pour plus d'informations, voir le plan de formation en annexe 4*

2- Les objectifs de collecte

L'objectif initial était fixé à 800 personnes enquêtées. Les objectifs par enquêteur·rices (nombre de questionnaires) étaient donc de 150 pour les enquêteur·rices et de 100 pour les référentes (chargées par ailleurs d'une mission de coordination).

La charge enquêteur a rapidement été jugée trop lourde (notamment au regard de la complexité du protocole et des obstacles liés au contexte COVID) mais le budget de l'enquête, tel que dimensionné dans le cadre de l'ANR, n'a pas permis de recruter davantage d'enquêteurs. Afin d'alléger la charge enquêteur, l'objectif a rapidement été révisé à 600 enquêté-es. L'équipe de terrain n'a été renforcée qu'en cours de collecte, avec une équipe de 8 personnes sur le terrain à partir de février 2021 (contre 6 de septembre 2020 à janvier 2021). Cela a conduit à organiser une seconde session de formation.

3- Organisation du terrain

- Su Wang et Ya-Han Chuang sont référentes et supervisent chacune jusqu'à quatre enquêteur-rices sur le terrain.
- Échanges réguliers entre les enquêteur-rices et Aurélie Santos (référente SES de l'enquête) pour : les lignes de rosters vides à supprimer, les rosters incomplets, les vérifications de doublons après le matching automatique.
- Échanges réguliers entre les enquêteur-rices et Isabelle Attané (responsable scientifique de l'enquête) pour : les refus, les envois de tickets cadeaux et consentement aux enquêté-es.
- Tous les lundis : les enquêteurs devaient mettre à jour leurs rosters incomplets en vue de l'échantillonnage du mardi.
- Le dernier mardi de chaque mois : point mensuel réunissant l'équipe recherche, les enquêteurs, la référente au SES.
- Bilan de fin de terrain le 30 juin 2021.

4- Outils de communication

a- Objectifs

Expliquer aux enquêté-es pourquoi il est important qu'ils participent à l'enquête et le rôle qu'ils ont à jouer pour nous aider à recruter de nouveaux enquêté-es au sein de leurs réseaux de sociabilité.

Expliquer aux enquêté-es ce qu'est l'Ined, les rassurer sur la préservation de leur anonymat et sur la confidentialité des données collectées, etc.

b- Supports développés

Réalisation des visuels : Isabelle Milan, service de la communication, Ined.

Alimentation du site web : Su Wang, chargée d'études UR04.

En collaboration avec le service de la communication de l'Ined, l'équipe recherche a conçu une E-PUB animée (voir annexe 5). Elle permet de présenter l'enquête de manière ludique et rapide. Elle a été diffusée sur le site de l'enquête et transmise par WeChat par les enquêteur-rices en cours de collecte.

Un site web de l'enquête a été créé, disponible en 2 langues : <https://ChIPRe.site.ined.fr> (en français) et <https://ChIPRe.site.ined.fr/en/home/> (en chinois)

Création d'un logo (en deux langues) utilisé sur tous les documents relatifs à l'enquête et notamment dans les messages WeChat (voir annexe 5).

Nous avons également conçu une plaquette d'information sur les associations qui accueillent des permanences en chinois (juridiques, médicales, etc.)

c- Moyens mobilisés en cours de collecte

Dès le tout début du terrain une enquêtee nous signale que WeChat génère un message d'alerte pour indiquer à nos interlocuteurs que notre compte pourrait être frauduleux ! Les enquêteurs et référentes pensent que c'est lié au fait que notre compte est trop récent et que nous y avons peu d'amis. Solutions déjà évoquées lors du test d'octobre : poster des articles sur l'avancée de l'enquête, poster des liens vers des articles de presse communautaires rédigés par les référentes (Huarenjie, Nouvelle Europe), mettre une photo de l'équipe des enquêteurs, montrer des graphes anonymisés d'avancée du réseau, nourrir le contenu au fil du terrain.

V- Bilan quantitatif de la collecte des données

1- Déroulement du terrain

a- Calendrier de la collecte et des confinements

Officiellement, le terrain de l'enquête ChIPRe s'est déroulé du **21 septembre 2020 au 30 juin 2021** (date du bilan de collecte avec les enquêteur·rices). Dans les faits, le terrain a démarré un peu plus tôt pour les référentes avec l'interview de 6 graines entre fin août et début septembre afin de dévoiler de premières personnes dans le réseau et être en mesure de confier des personnes à enquêter aux enquêteurs à l'issue de leur formation (14 au 16 septembre). La collecte s'est également un peu prolongée au-delà du 30 juin avec une reprise de l'exploitation des personnes à enquêter par les référentes, à la fin des contrats des enquêteur·rices. Cette reprise des cas « sans issue finale » au 30 juin n'a abouti à la réalisation que d'un seul questionnaire début juillet. Le terrain a donc duré entre 9 et 10 mois.

Le terrain a été marqué par l'instauration du deuxième confinement, obligeant la collecte à basculer du face-à-face vers le distanciel. La fin du confinement n'a toutefois pas permis la reprise du face-à-face : le couvre-feu empêchait de fixer des RDV en soirée, et l'Ined a pris des mesures restrictives, suspendant les entretiens en face à face au début du confinement (à partir du 01/11/2020, le personnel Ined bascule en télétravail à temps plein). La reprise du terrain s'est faite au moment où le retour sur place à l'Ined pour 1 jour par semaine a été toléré, fin mai 2021.

Rappel des différentes mesures et restrictions mises en place sur la période de collecte

29 octobre 2020 – 14 décembre 2020 : Deuxième confinement national

28 novembre 2020 : Allègement du confinement :

- *L'attestation de déplacement reste en vigueur mais les déplacements brefs (promenade, loisirs...) sont autorisés dans un rayon de 20 km (contre 1 km auparavant) et pour une durée de 3h maximum (contre seulement 1h lors du reconfinement).*
- *Les commerces dits « non-essentiels » vont également rouvrir avec un protocole sanitaire renforcé avec une limite d'un client pour 8 mètres carrés, seuls les restaurants, bars, cafés et le monde de la culture restent porte close.*

15 décembre 2020 : fin du confinement, mise en place du couvre-feu (20h-6h)

16 janvier 2021 : Couvre-feu avancée à 18h

À partir du 20 janvier 2021 : réouverture des restaurants, réouverture totale des lycées

20 mars 2021 : Couvre-feu repoussé à 19h

19 mai 2021 : Couvre-feu repoussé à 21h

9 juin 2021 : Couvre-feu repoussé à 23h, assouplissement du télétravail, évolution des conditions d'accueil des commerces et lieux culturels et de loisirs

17 juin 2021 : Levée de l'obligation du port du masque en extérieur sauf dans les situations de regroupements de personnes (marchés, stades, files d'attente...)

20 juin 2021 : Levée du couvre-feu

30 juin 2021 : Fin des limites de jauges dans les lieux recevant du public.

	Sept 2020	Oct 2020	Nov 2020	Déc. 2020	Janv. 2021	Fév. 2021	Mars 2021	Avril 2021	Mai 2021	Juin 2021
1	Ma	J	D	Ma	V	L	L	J	S	Ma
2	Me Ref	V	L	Me	S	Ma	Ma	V	D	Me
3	J	S	Ma	J	D	Me	Me	S	L	J
4	V	D	Me	V	L	J	J	D	Ma	V
5	S	L	J	S	Ma	V	V	L	Me	S
6	D	Ma	V	D	Me	S	S	Ma	J	D
7	L	Me	S	L	J	D	D	Me	V	L
8	Ma	J	D	Ma	V	L 2	L	J	S	Ma
9	Me	V	L	Me	S	Ma	Ma	V	D	Me
10	J	S	Ma	J	D	Me	Me	S	L	J
11	V	D	Me	V	L	J	J	D	Ma	V
12	S	L	J	S	Ma	V	V	L	Me	S
13	D	Ma	V	D	Me	S	S	Ma	J	D
14	L 1	Me	S	L	J	D	D	Me	V	L
15	Ma	J	D	Ma	V	L	L	J	S	Ma
16	Me	V	L	Me	S	Ma	Ma	V	D	Me
17	J	S	Ma	J	D	Me	Me	S	L	J
18	V	D	Me	V	L	J	J	D	Ma	V
19	S	L	J	S	Ma	V	V	L	Me	S
20	D	Ma	V	D	Me	S	S	Ma	J	D
21	L -DEBUT-	Me	S	L	J	D	D	Me	V	L
22	Ma	J	D	Ma	V	L	L	J	S	Ma
23	Me	V	L	Me	S	Ma	Ma	V	D	Me
24	J	S	Ma	J	D	Me	Me	S	L	J
25	V	D	Me	V	L	J	J	D	Ma	V
26	S	L	J	S	Ma	V	V	L	Me	S
27	D	Ma	V	D	Me	S	S	Ma	J	D
28	L	Me	S	L	J	D	D	Me	V	L
29	Ma	J	D	Ma	V		L	J	S	Ma
30	Me	V	L	Me	S		Ma	V	D	Me -FIN-
31		S		J	D		Me		L	

	Formations référentes et enquêteurs (1 et 2)
	Dates de collecte
	Confinement
	Couvre-feu 18h
	Couvre-feu 19h
	Couvre-feu 20h
	Couvre-feu 21h
	Couvre-feu 23h

b- Calendrier des tirages

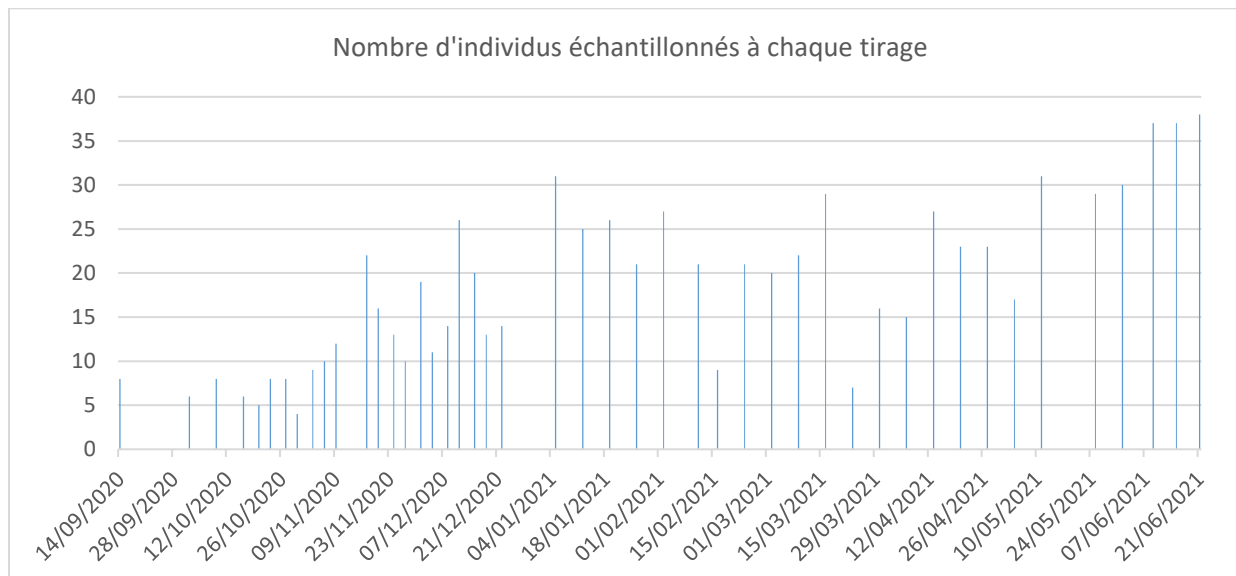
La méthode NSM nécessite la réalisation de nombreux tirages au fil du terrain, au fur et à mesure que l'on dévoile le réseau. Durant les quelques neuf mois de terrain, près de 50 tirages ont été réalisés (en bleu dans le calendrier ci-dessous).

	Oct 2020	Nov 2020	Déc. 2020	Janv. 2021	Fév. 2021	Mars 2021	Avril 2021	Mai 2021	Juin 2021
1	J	D	Ma	V	L	L	J	S	Ma
2	V	L	Me	S	Ma	Ma	V	D	Me
3	S	Ma	J	D	Me	Me	S	L	J
4	D	Me	V	L	J	J	D	Ma	V
5	L	J	S	Ma	V	V	L	Me	S
6	Ma	V	D	Me	S	S	Ma	J	D
7	Me	S	L	J	D	D	Me	V	L
8	J	D	Ma	V	L	L	J	S	Ma
9	V	L	Me	S	Ma	Ma	V	D	Me
10	S	Ma	J	D	Me	Me	S	L	J
11	D	Me	V	L	J	J	D	Ma	V
12	L	J	S	Ma	V	V	L	Me	S
13	Ma	V	D	Me	S	S	Ma	J	D
14	Me	S	L	J	D	D	Me	V	L
15	J	D	Ma	V	L	L	J	S	Ma
16	V	L	Me	S	Ma	Ma	V	D	Me
17	S	Ma	J	D	Me	Me	S	L	J
18	D	Me	V	L	J	J	D	Ma	V
19	L	J	S	Ma	V	V	L	Me	S
20	Ma	V	D	Me	S	S	Ma	J	D
21	Me	S	L	J	D	D	Me	V	L
22	J	D	Ma	V	L	L	J	S	Ma
23	V	L	Me	S	Ma	Ma	V	D	Me
24	S	Ma	J	D	Me	Me	S	L	J
25	D	Me	V	L	J	J	D	Ma	V
26	L	J	S	Ma	V	V	L	Me	S
27	Ma	V	D	Me	S	S	Ma	J	D
28	Me	S	L	J	D	D	Me	V	L
29	J	D	Ma	V		L	J	S	Ma
30	V	L	Me	S		Ma	V	D	Me
31	S		J	D		Me		L	

Sans recommandations de la part de l'équipe conceptrice de la méthode, nous avons démarré le terrain sans savoir à quelle fréquence nous devions réaliser le matching et l'échantillonnage de nouveaux individus.

Dans un premier temps, nous avons donc tâtonné et sommes partis sur un rythme de deux tirages par semaine. Cependant, nous avons rapidement fait le constat que cette tâche prenait du temps, entre le matching, le contrôle manuel des doublons potentiels, l'échantillonnage et l'affectation des nouvelles personnes à enquêteur aux enquêteurs. De plus, la collecte de lignes complètes de roster prenant du temps (voir graphique *Durée pour obtenir un roster complet*, plus bas), les individus nouvellement apparus dans le réseau ne sont pas tout de suite candidats à l'échantillon (car lignes incomplètes). Avec des échantillonnages rapprochés, on avait des viviers de taille très restreinte. Pour toutes ces raisons, nous avons décidé, à compter de janvier, de passer à un tirage par semaine, à jour fixe (le mardi), afin que les enquêteurs tentent de collecter avant le mardi matin les informations manquantes des rosters collectés (une fois les informations complétées, la ligne de roster est validée, et passe au matching/échantillonnage).

Lorsque nous procédions à deux tirages par semaine, ces derniers étaient de taille relativement restreinte (moins de 15, voire moins de 10). À compter de janvier, avec le passage à un seul tirage par semaine, les échantillons tirés étaient plus importants (24 individus en moyenne).



Au final, sur les 1523 individus identifiés dans le réseau, on compte :

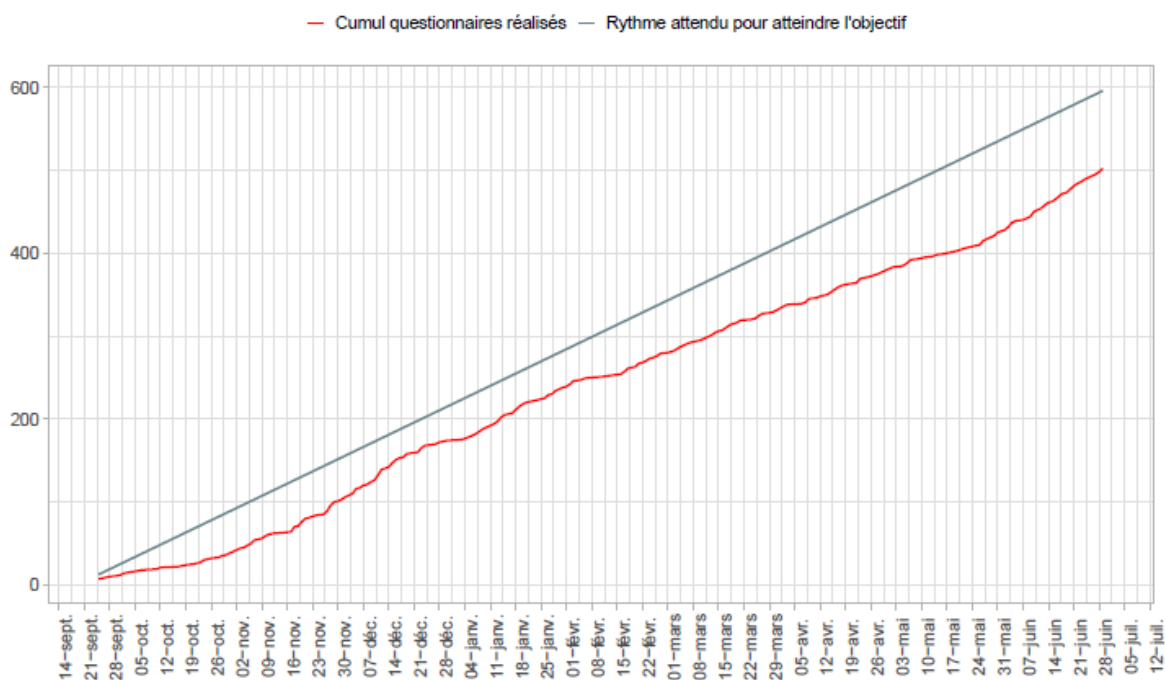
- 16 graines (toutes enquêtées)
- 844 individus échantillonnés.

Le taux de sondage du réseau dévoilé grâce à l'enquête est donc de 56,5% (16+844/1523).

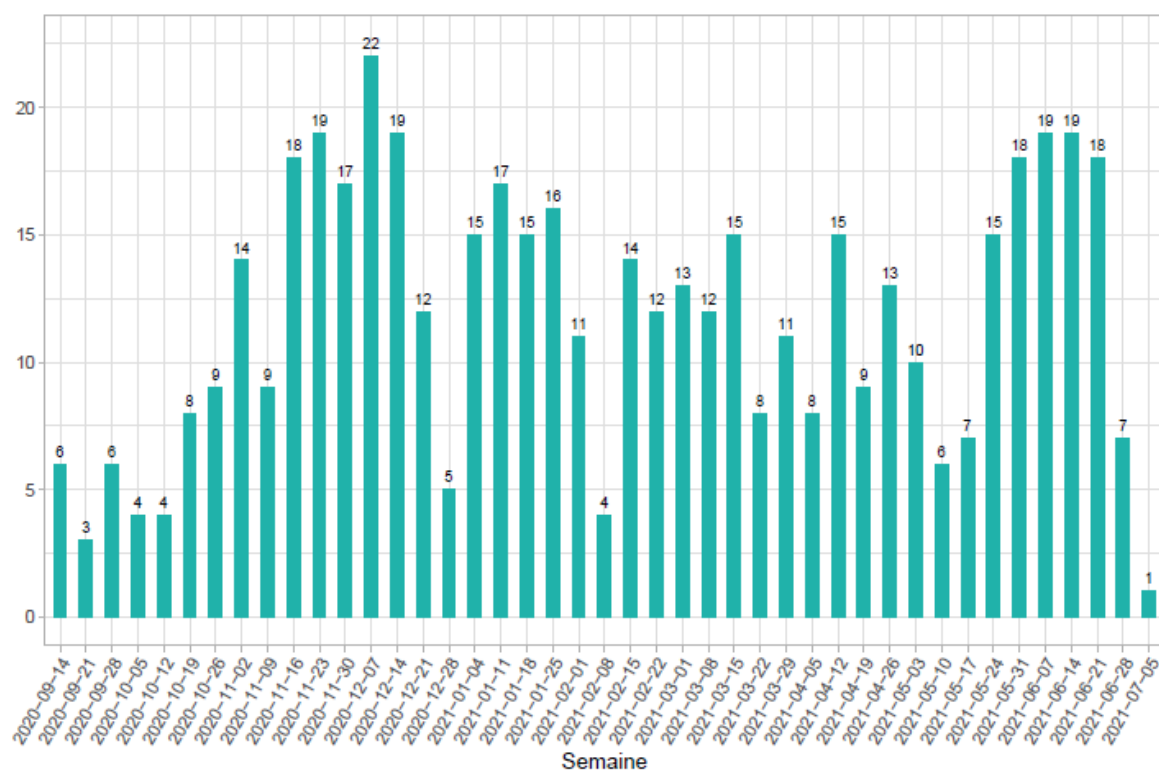
c- Rythme de collecte

Rythme de collecte

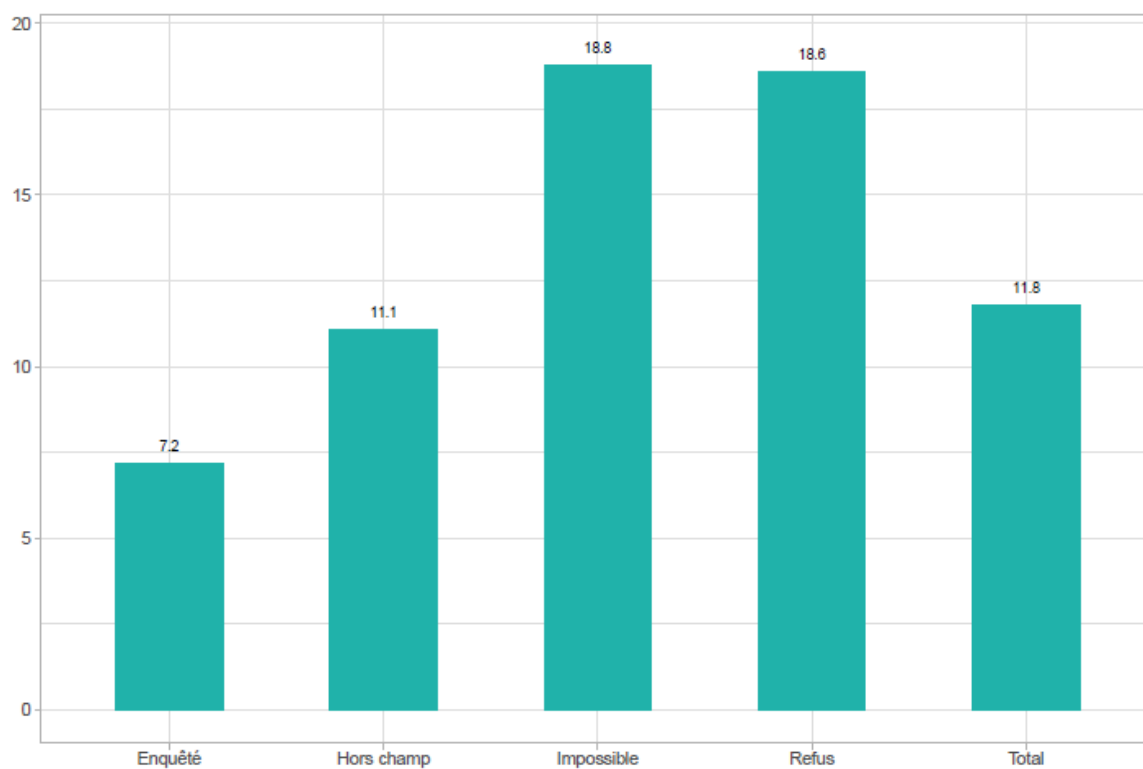
Totalité du terrain



Nombre de questionnaires collectés par semaine

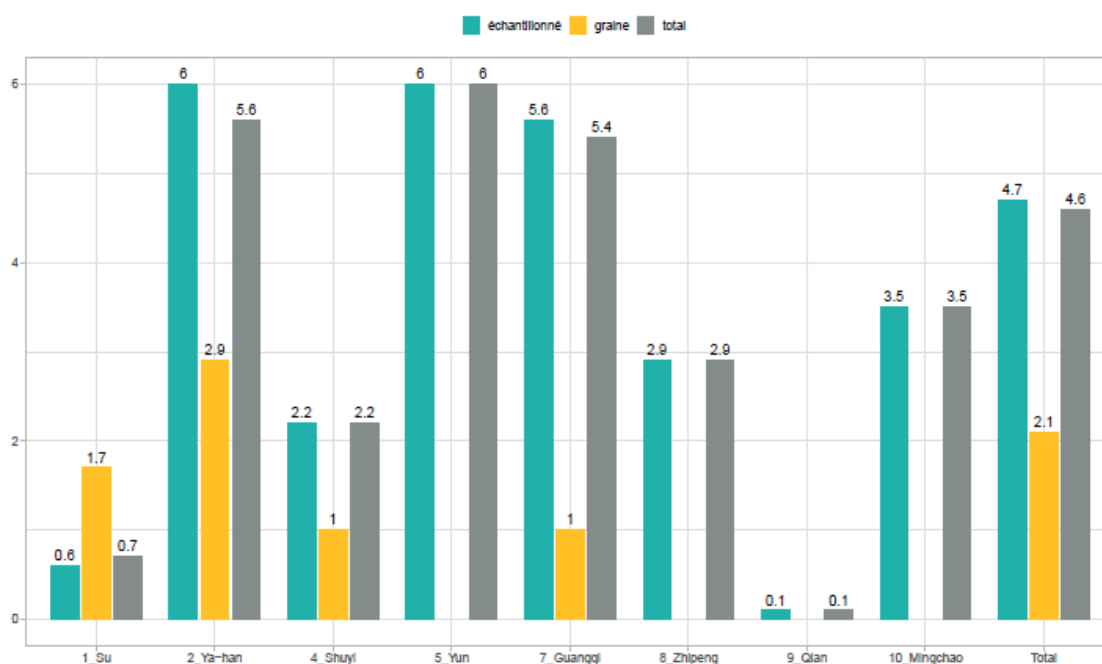


Durée moyenne d'exploitation des échantillons (en nombre de jours) en fonction de l'issue



Durée pour obtenir un roster complet (non vide)

Nombre de jours écoulés entre l'interview du citant et le dernier passage d'une ligne du roster à la modalité "1. Complète" (donc date à laquelle un roster est estimé complet)



2- Bilan en fin de terrain

a- Chiffres clés

Parmi les 844 individus échantillonnés (non graines), on compte :

- 486 répondants, ce qui correspond à un taux de réponse brut de 57,6%
- 328 refus, soit un taux de refus brut de 38,9%
- 24 personnes hors champ
- 6 personnes impossibles à enquêter.

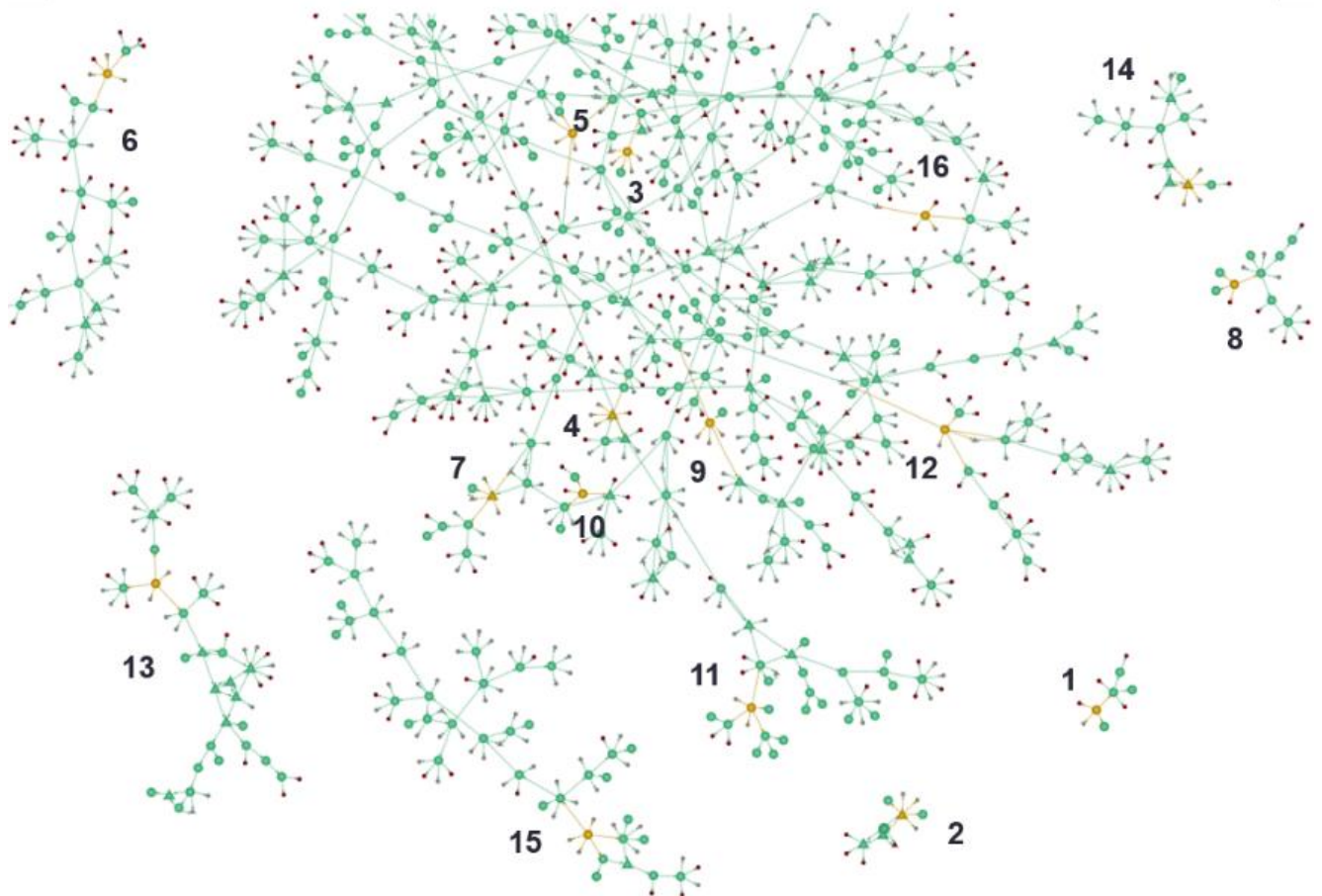
Pour un calcul plus juste des taux de réponse et taux de refus, il convient en suivant les recommandations de l'AAPOR d'exclure du dénominateur les personnes identifiées comme étant hors champ, le taux de réponse est donc de 59,3% et le taux de refus de 40%.

Nombre d'enquêtés	501
Graines	16
Échantillonnés	485
Nombre de citations	1698
Nombre d'individus dans le réseau	1523
Nombre d'individus cités 1 seule fois	1357
Nombre d'individus cités 2 fois	133
Nombre d'individus cités 3 fois	18
Nombre d'individus cités 4 fois	3

Parmi les enquêtés :

- 71% ont participé par téléphone
- 19% ont rempli le questionnaire en visio avec l'enquêteur·trice
- 10% en face-à-face

b- État final du réseau (au 30 juin 2021)

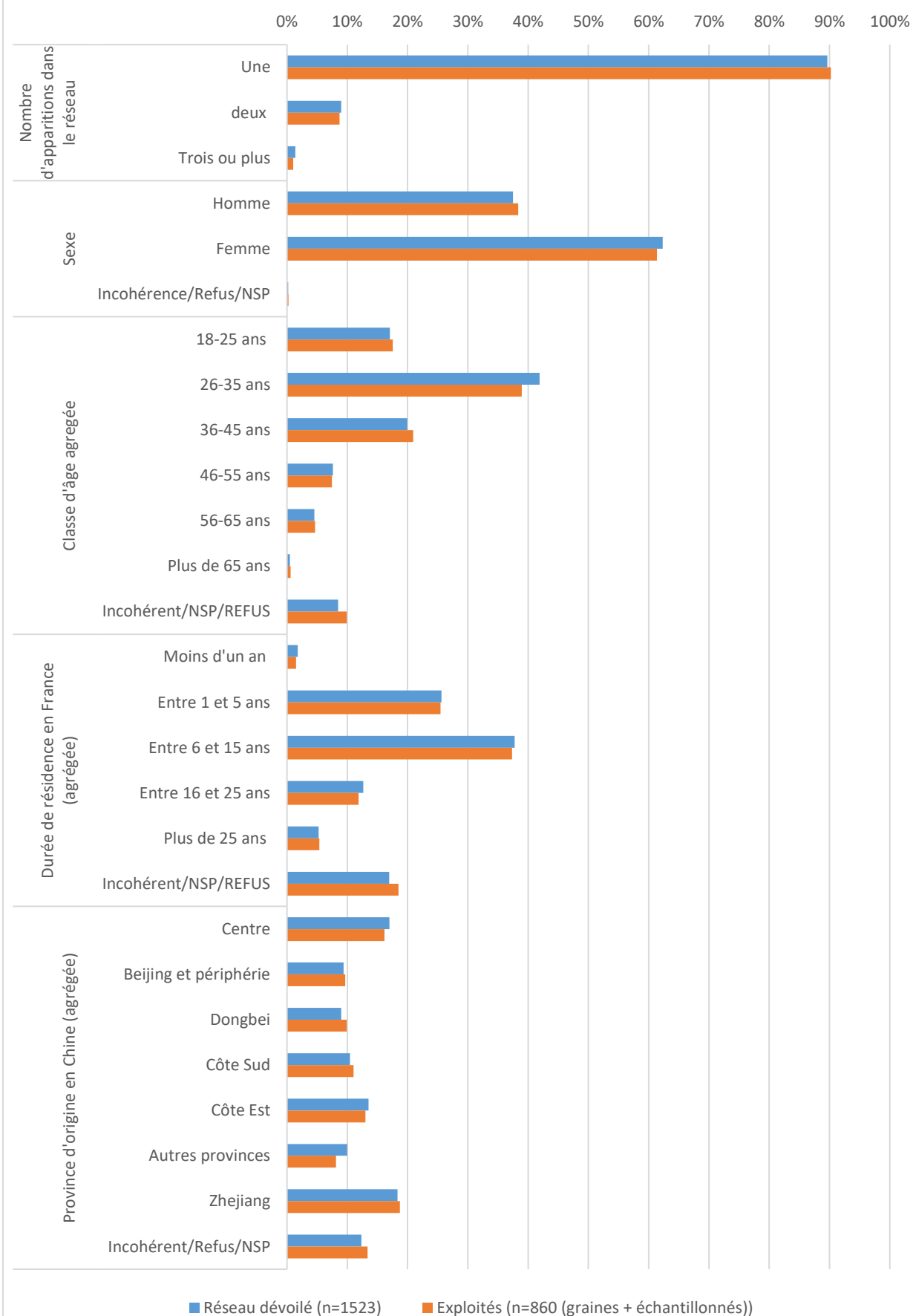


- En jaune : les graines (et leur numéro)
- En rouge : les refus, impossibles ou hors-champs
- En triangle : les individus faisant l'objet de citations multiples

c- Profils des échantillonnés par rapport au réseau dévoilé

Sur le nombre d'apparitions dans le réseau (nombre de fois où l'individu a été cité), le sexe, l'âge, la durée de résidence en France et la province d'origine en Chine, on observe une très grande similarité entre le réseau dévoilé (n=1523) et l'échantillon d'individus exploités (n=860). L'échantillon exploité est donc représentatif du réseau sur ces critères. Le processus de sélection/d'échantillonnage des individus ne distord pas l'échantillon, ce qui est une bonne chose.

Comparaison entre réseau dévoilé et échantillon exploité



d- Profils des échantillonnés selon l'issue

Les répondants et refusants représentent plus de 95% des individus exploités. Les impossibles et hors cible représentent donc moins de 5%. Dans cette partie, nous nous attacherons donc à décrire les deux premières catégories.

Pour cela, nous avons modélisé la probabilité d'avoir participé et celle d'avoir refusé en fonction :

- Des caractéristiques sociodémographiques du cité (renseignées par le/les citant-s) :
 - o Sexe
 - o Classe d'âge
 - o Durée de résidence en France
 - o Province d'origine en Chine
- Des caractéristiques du/des lien-s avec le/les citant-s :
 - o Nombre de citations
 - o Fréquence des contacts entre le citant et le cité (indiquée par le citant)
 - o Caractérisation de la relation entre le citant et le cité (indiquée par le citant)
- Des coordonnées fournies par le citant pour joindre le cité

Profils des répondants

Parmi les caractéristiques prises en compte dans la modélisation, on note assez peu d'effets significatifs (points verts ou rouges) par rapport à la situation de référence.

Les caractéristiques associées positivement à la participation sont :

- Le fait d'avoir été cité plusieurs fois
- Le fait d'être en France depuis plus de 10 ans (ou que le citant n'ait pas su indiquer la durée de résidence en France)

Les caractéristiques associées négativement à la participation sont :

- Le fait d'être âgé de 36 à 45 ans (plutôt que moins de 25 ans)
- Le fait que le citant n'ait pas indiqué le compte WeChat du cité
- Le fait de le citant n'ait pas indiqué le numéro de téléphone du cité

Bien que non significatives, on observe par ailleurs les tendances suivantes :

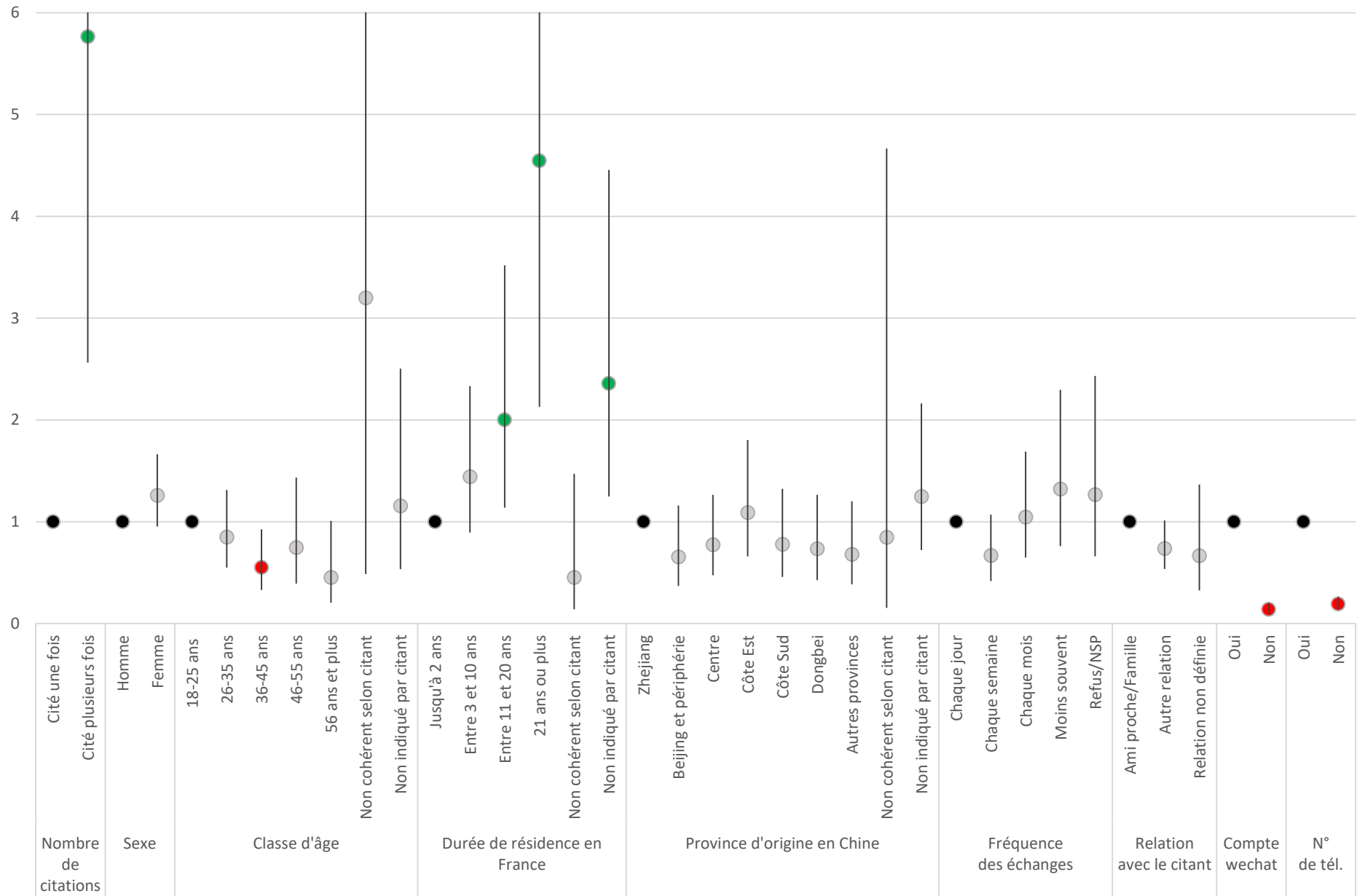
- Moindre participation des hommes
- Décroissance de la probabilité de répondre avec l'âge
- Moindre probabilité de réponse chez les personnes arrivées récemment en France
- Plus de participation lorsque le citant est un ami proche ou un membre de la famille du cité.

Profils des refusants

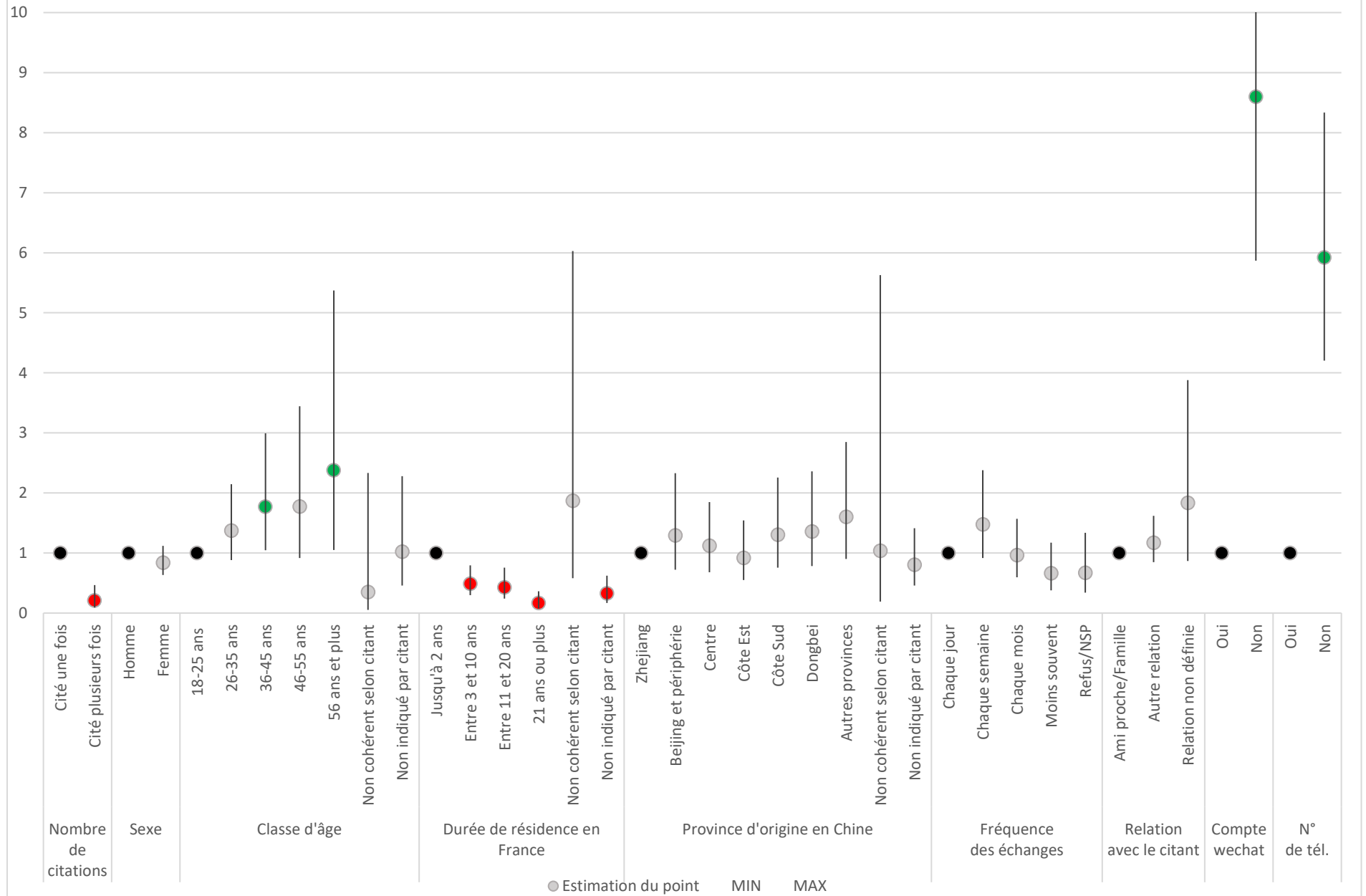
Les caractéristiques associées à une propension plus élevée à refuser sont :

- Le fait d'être âgé entre 36 et 45 ans ou de plus de 55 ans
- L'absence de compte WeChat
- L'absence de numéro de téléphone.

Modélisation de la participation - Rapports de côtes



Modélisation des refus - Rapports de côtes



À l'inverse, les caractéristiques associées à une propension moindre de refuser sont :

- Le fait d'avoir été cité plusieurs fois
- Le fait d'être arrivé en France depuis plus de deux ans (par rapport au fait d'être arrivé depuis 2 ans ou moins)

⇒ *Pour plus d'informations, voir le document « Analyses menées sur les rosters » (annexe 6)*

VI- Les difficultés du terrain (bilan qualitatif)

La collecte des données de l'enquête ChiPre s'est déroulée de septembre 2020 à juin 2021 et a mobilisé au total 10 enquêteur·rices et une équipe de 5 gestionnaires de collecte (informaticien, statisticien·nes, ingénieure d'études).

Cette partie vise à retracer les principaux obstacles rencontrés sur le terrain, et solutions mises en œuvre au fil de la collecte pour tenter d'y remédier. Elle s'appuie sur de multiples documents : les CR des points mensuels avec les référentes et enquêteur·rices, les CR des réunions de l'équipe recherche et de la référente du SES menées à rythme régulier durant toute la collecte, les questionnaires remplis par les enquêteur·rices en fin de terrain, le bilan final du 30 juin.

1- Les implications d'une enquête par réseau

a- La difficile transmission des réseaux entre enquêteur·rices

Même si elles sont couteuses, nous avons rencontré trois cas où les « transmissions de réseaux » ou « ruptures de filiation » entre enquêteur·rices se sont révélées indispensables :

1) Transition référentes → enquêteur·rices en septembre 2020

Les référentes connaissent bien les graines et les ont enquêtées début septembre 2020, à ce stade elles sont donc en position privilégiée pour négocier les RDV et réaliser les prises de contact avec les graines et leurs cités pour les nouveaux enquêteurs. La mise en contact ne se fait pas facilement car le cité a déjà du mal à accepter d'ajouter le compte WeChat de la personne qui a enquêté la graine, donc cela peut être difficile d'accepter 2 personnes nouvelles d'un coup.

2) Transition entre enquêteur·rices en cours de collecte en fonction de leur charge de travail

Ce choix a fait l'objet de nombreuses réclamations de la part des enquêteurs (point mensuel du 27/10) : il faut beaucoup d'interactions entre les cités et l'enquêteur pour faire accepter un autre enquêteur, c'est très peu incitatif, cela risque de peser sur le nombre de citations.

Afin d'expliquer aux enquêteurs que nous étions obligés d'en passer par là en début de collecte, Géraldine Charrance a fait une simulation : les performances étant très inégales, si on avait respecté strictement toutes les filiations, 40% des échantillonnés seraient allés à une enquêtrice uniquement et plusieurs enquêteurs n'auraient jamais rien reçu à l'issue des tirages → on est obligé parfois de « rompre » les filiations et redistribuer les échantillonnés entre enquêteurs. De plus, certains se plaignent de rester coincés dans des réseaux précaires difficiles : nous rappelons à cette occasion que les ruptures de filiation permettent de faire varier les profils enquêtés.

Cependant nous avons entendu les retours de terrain et nous avons décidé de laisser complètement tomber l'objectif d'équité entre eux, et de simplement suivre les filiations au maximum à l'issue des tirages.

3) Transition entre les anciens et les nouveaux enquêteurs en février 2021

Nous avons essayé d'anticiper au maximum cette étape en recommandant aux enquêteurs nous quittant de préparer leurs derniers enquêtes à ces transferts, mais cela n'a pas toujours été bien fait. Nous avons décidé, à l'issue des tirages réalisés à partir de mi-février 2021, de systématiquement rediriger les échantillons issus des réseaux des enquêteurs sortants vers les deux référentes, plus aguerries aux négociations. Elles ont donc récupéré chacune un des téléphones pour assurer la suite du terrain au sein des filiations amorcées.

b- La négociation des rosters

Ce sujet est de loin celui qui a le plus préoccupé l'équipe de terrain et de gestion du terrain. Une grande partie des échanges lors des points mensuels avait pour objet le partage d'expérience entre enquêteur·rices, à la recherche de solutions et de soutien moral.

Les difficultés de négociation

Tous les enquêteur·rices ont fait état de la difficulté à obtenir des citations de roster au-delà de 2 ou 3 citations. Cela demande énormément d'effort de collecter des rosters complets.

Il est très rarement possible d'obtenir un roster complet en fin de questionnaire : il faut revenir vers le citant une fois qu'il a contacté ses amis pour leur demander leur accord de transmettre leurs coordonnées. En général : le citant en parle à ses amis, puis transfère le WeChat Card du cité à l'enquêteur s'ils acceptent. S'il ne le fait pas : c'est ce que l'on a appelé le blocage « indépassable » du citant. Le second blocage indépassable (lorsque l'on n'a pas le téléphone) est le refus du cité d'accepter l'invitation WeChat. Ces blocages ont beaucoup nourri la frustration des enquêteur·rices.

Il est difficile de demander le WeChat ET le téléphone, il faut souvent se contenter de l'un des deux, pourtant lorsque l'invitation WeChat n'est pas acceptée il peut être salvateur de bénéficier d'un numéro de téléphone.

La difficulté croît :

- Avec la distance aux graines : Plus on s'éloigne des graines, plus le lien de confiance est ténu, plus cela renforce les difficultés de négociation
- Avec certains profils : les réseaux les plus précaires, les sans papier, la communauté Wenzhou.
- Lorsque des anciens rosters ressurgissent à l'échantillonnage plusieurs semaines/mois après leur collecte
- Avec le distanciel, quel que soit le profil !

Techniques de négociation

Technique « grand roster »: lors du second point mensuel (27/10/2021), une enquêtrice nous fait part de sa technique: elle insiste pour collecter le maximum de citation à la fin du questionnaire. Souvent il était difficile d'obtenir les coordonnées (tel ou WeChat) au-delà du 2^e ou 3^e contact, mais au moins, elle avait une liste plus conséquente de cités et plus d'accroche pour revenir ensuite vers le citant pour négocier le reste des coordonnées. Elle parvient à cela en promettant qu'elle ne va pas les contacter sans leur accord, et qu'elle revient vers eux si les cités sont échantillonnées. Cela rassure et incite à la citation.



L'avantage est d'avoir une meilleure description du réseau. Et un roster rapidement échantillonnable (même si tous les contacts ne sont pas renseignés) ce qui ne pèse pas trop sur le rythme de collecte.

— L'inconvénient est d'avoir un taux de refus plus important ensuite.

Technique « petits rosters » : attendre l'accord des citants pour compléter la moindre ligne de roster, donc collecter des rosters plus petits et composés des membres du réseau du citant dont il a la quasi-certitude qu'ils seront d'accord avec le principe de l'enquête.

+ L'avantage est d'avoir souvent un taux de refus plus bas dans ces filiations.

— L'inconvénient est l'allongement certain du rythme de collecte car il faut attendre que le citant sollicite ses amis avant de pouvoir remplir ne serait ce qu'une ligne de roster. De plus cela alourdit la charge mentale du citant : il endosse une charge de négociation qui reviendrait normalement à l'enquêteur. Enfin, ne citer que des « candidat idéaux » biaise fortement l'image du réseau de sociabilité des enquêtés.

Bien sûr dans les faits, il s'agit plus souvent d'un mix des 2 méthodes, utilisées au cas par cas.

Argumentaires déployés

Dans les questionnaires de bilan de terrain, tous mentionnent à l'unanimité que les tickets cadeaux ont été indispensables à la négociation des rosters. L'argument des chèques cadeaux est plus porteur chez les plus précaires mais reste un atout pour faire basculer des négociations hésitantes chez d'autres profils également.

Une enquêtrice mentionne que les plus précaires ne sont pas nécessairement les plus réfractaires. La fragilité de leur statut fait que l'on a plus de moyens de rendre des petits services pour gagner leur confiance, tandis que l'on n'a rien à offrir aux CSP+ !

L'argument de la discrimination mobilise assez peu les plus dotés, qui se sentent plus protégées des discriminations et moins touchés.

Chaque public a besoin d'un argumentaire adapté. Les messages WeChat de contact doivent donc être personnalisés en fonction de ce que l'on sait du cité.

Plusieurs enquêteur·rices ont envoyé leurs vœux sur WeChat lors du nouvel an chinois en février 2021, ce qui a été suivi de peu d'effet.

Une enquêtrice, à court d'arguments, a essayé la stratégie de compassion : « On a vraiment du mal à trouver des enquêtés » pour que les gens citent des contacts (20/04/2021).

Tous redoublent d'imagination : un procès a eu lieu en mars 2021 au sujet de cyber harcèlement contre des chinois. Les enquêteur·rices en parlent pour intéresser les enquêtés et montrer que c'est important de rendre le sujet des discriminations plus visible. Une enquêtrice a fait un post sur son WeChat et cela a débloqué un lien vers des cités car le citant a liké le post !

Aider le citant à convaincre les cités

Plusieurs enquêtrices mentionnent que les cités ne sont parfois pas prévenus quand elles les contactent, ce qui rend l'échange plus difficile.

Il peut être couteux pour un enquêté de prendre contact avec ses cités, surtout s'il en a cité 6 ! Les enquêteur·rices proposent de l'aide avec des messages « tout fait ». Pour aider les enquêtés à parler aux cités, une enquêtrice leur propose de faire des messages audios, mais ils utilisent peu ce moyen de communication et préfèrent l'écrit.

Principales réticences rencontrées

« Pas envie de déranger leurs amis » est la raison numéro 1. Beaucoup considèrent que c'est un échange de service de répondre au questionnaire, donc c'est une dette envers ses amis (s'ils répondent parce qu'on les a cités). Ils ne veulent pas s'embêter avec ça et préfèrent ne pas les citer.

Le fait d'être cité par quelqu'un ne garantit pas l'effet « recommandation par les pairs » :

- Parfois le lien est asymétrique, non réciproque, donc la citation a peu de poids
- Parfois le lien est fragile, donc on ne peut pas s'appuyer sur le citant
- Exemple de cas d'asymétrie : une enquêtée était très déçue du refus de ses amis, car elle découvre à cette occasion que ses amis ne sont pas prêts à faire ce qu'elle a fait pour sa citante ! Elle a accepté pour ne pas faire « perdre la face » de sa citante, or ses amis ne tiennent pas à en faire autant pour elle.
- Une citée n'a pas voulu répondre car, certes elle connaissait le citant, mais l'enquêtrice ne faisant pas partie du réseau de connaissance du citant, elle n'a pas eu suffisamment confiance pour répondre.

Le fait de montrer que l'on ne saisit pas le prénom en entier dans le roster ne constitue pas un levier suffisant pour la confiance. Cela en rassure certains à la marge.

Pas d'entretien de leur réseau de sociabilité pendant le Covid, donc encore plus difficile de les déranger pour une enquête.

Motif de réticence plus rarement exprimé : « est-ce que c'est politique ? »

c- Des réseaux précaires qui nécessitent une implication importante

Les cités de la graine 1 constituent un réseau complexe, vivant des situations de violence et d'extrême précarité (violences conjugales, hospitalisation, démêlés avec la police, etc.) L'investissement en temps et ressources sociales mobilisés pour enquêter ce réseau sont considérables. Les logiques de don-contre don à l'œuvre nécessitent une implication qui déborde de loin la tâche de passation de questionnaire. Dans ce contexte il n'est pas possible de transférer ce réseau vers les nouveaux enquêteurs.

En octobre, les enquêteurs ayant mené des entretiens avec des sans-papier font part des situations sociales et parcours migratoires très difficiles qu'ils ont recueillis. Le contexte de passation est complexe : par exemple, entretien au domicile d'un enquêté, qu'il partage avec 7 autres sans papier, qui rend difficile la passation dans le calme avec la confidentialité requise. Une propriétaire de logement a un jour expulsé l'enquêteur en raison de craintes liées au Covid (mais sans doute aussi par rapport à la sous-location du logement) qui a terminé son questionnaire dans un couloir.

Les attentes de ces enquêté-es sont aussi difficiles à gérer : il-elles pensent que l'enquêteur est travailleur social, ou qu'il a des connaissances de droit, et posent beaucoup de questions qui dépassent le rôle de l'enquêteur (administratif, banque alimentaire etc.).

De plus, le temps passé à essayer de convaincre ces réseaux n'est pas récompensé en conversions en enquête en proportion.

- ⇒ Il est important de noter que, conformément à ce qui a déjà été observé lors de la *formative research* et du test de protocole d'octobre, les conditions d'enquête auprès des publics les plus précaires nécessitent un fort investissement de la part des enquêteurs (logique de don/contre-

don, temps passé sur place pour créer de la confiance, réception des récits de parcours heurtés ou témoignages de violence) et nécessitent d'être anticipées dans les formes de rémunération mises en place et dans l'explicitation des tâches à réaliser : il s'agit certes d'une enquête quantitative, mais qui comporte une dimension qualitative où l'exigence d'implication sur le terrain est forte (ce qui rend par ailleurs le temps passé sur le terrain d'autant plus difficile à quantifier). Une autre piste serait de développer de véritables partenariats avec des associations venant en aide aux sans papier et proposant des services en langue chinoise.

d- La gestion des graines complémentaires

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 (dont la forte présence dans le recensement se reflétait mal dans notre réseau), et les personnes les plus précaires ont fait l'objet d'une attention particulière à cet égard. Entre mi-octobre 2020 et début avril 2021, les graines 9 à 16 ont ainsi été ajoutées (voir détail en annexe 1).

Les référentes doivent chercher d'autres candidats à l'enquête dans leur réseau ou dans le réseau créé lors de la *formative research*, alors que leur charge d'enquête et leur rôle de « premier interlocuteur » pour les difficultés rencontrées par les enquêteur·rices sur le terrain les occupent déjà beaucoup. Certaines enquêtrices ont même été sollicitées pour aider au recrutement des graines complémentaires.

Ces ajouts ont suscité beaucoup de discussions entre l'équipe recherche et le SES. En l'absence de recommandations existantes, il a parfois été difficile de trancher sur ces ajouts.

➤ Arguments pour :

- C'est une piste pour relancer une dynamique positive de l'enquête et sortir du cercle du faible nombre de citations.
- Si on atteint 600 enquêtés, une vingtaine de graines au maximum représente seulement 3% des enquêtés.
- Ajouter des graines pour l'arrivée des nouveaux enquêteurs en février 2021 semble logique : cela leur permet de lancer leur propre filiation sans souffrir d'une transmission de réseau de précédents enquêteurs vers les nouveaux (que l'on sait être très couteuse en énergie et temps de négociation).

➤ Arguments contre :

- Toutes les graines n'auront pas le même temps d'exploitation et la même exposition au tirage, il est difficile de prévoir les conséquences induites au niveau de la pondération.
- Les réseaux les moins accessibles le sont pour des raisons que l'ajout de graines ne comblera pas forcément : la difficulté à intégrer des réseaux complexes comme les sans papier et les Wenzhou provient surtout de l'impossibilité d'enquêter en face à face et de passer du temps avec les enquêtés pour créer un lien de confiance suffisamment solide.
- Les quartiers à forte concentration de chinois et mal représentés dans l'échantillon (cf volonté d'ajouter la G14) sont caractérisés notamment par deux types de profils : des travailleurs précaires (souvent sans-papiers) et des commerçants de Wenzhou. Donc le critère de la localisation « cache » d'autres caractéristiques sociodémographiques. Or, ces deux réseaux pour lesquels nous avons déjà des graines s'exposent à un très fort taux de refus. En témoignent les filiations difficiles des graines 2, 6 (Aubervilliers), 8 (Pantin) et 11 (Bagnolet).

2- Des difficultés spécifiques à la méthode NSM

a- Un algorithme qui nécessite de nombreux ajustements en cours de collecte

La première difficulté à laquelle s'est trouvée confrontée l'équipe gestionnaire a été l'élaboration de règles encadrant les échantillonnages. La méthode ayant très peu été testée sur le terrain, nous ne disposons pas de recommandations précises sur ces échantillonnages successifs.

À quelle fréquence réaliser les 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 terrain et du réseau au démarrage de la collecte mais s'est rapidement avérée trop chronophage et nous avons alors décidé de ne réaliser qu'un seul tirage plus conséquent par semaine (et exceptionnellement de gros tirages à la veille de périodes de congés).

De quelle taille doivent être les tirages ?

Pour déterminer le nombre d'individus tirés à chaque phase d'échantillonnage, nous mobilisons les paramètres suivants :

- La charge d'enquête restante pour chaque enquêteur·rice ;
- Leur charge de relance et d'identification des rosters incomplets (un enquêté donne des citations dans son roster mais souhaite d'abord demander à ses amis leur accord pour transmettre leurs coordonnées, il faut alors le relancer régulièrement pour obtenir ces informations) ;
- La taille du réseau au moment du tirage par rapport à sa taille lors du précédent.

Après diverses expérimentations, nous avons décidé de produire des tirages de taille modulable sans jamais dépasser la moitié des nouveaux arrivés dans le réseau depuis le précédent tirage.

L'inconvénient de cette méthode est que la taille des tirages ne peut pas être anticipée. Il faut d'abord identifier les doublons afin de connaître le nombre de nouveaux individus ajoutés dans le roster total.

Comment les répartir entre enquêteurs ?

Nous avons démarré le terrain avec un objectif fort 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 il était en contradiction avec le respect des « filiations » sur le terrain. Les enquêteur·rices souhaitaient suivre « leur » réseau, c'est-à-dire enquêter les cités de « leurs » enquêté·es (si échantillonné·es). Or, une répartition équitable d'échantillonnés entre eux impliquait des « transfert de filiation », qui se sont révélés beaucoup trop coûteux sur le terrain. 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.

Des modifications de l'algorithme 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.

Sans entrer dans le détail de toutes les modifications mises en œuvre (voir annexe 7), 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 autres, 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 II) afin d'avoir plus de chances de l'atteindre et d'expérimenter la « sortie » du mode *Search*. Ce seuil a été tiré de l'article fondateur de la méthode NSM, dans lequel diverses versions de ce seuil sont présentées (Mouw, 2012).

b- Un algorithme d'échantillonnage qui produit des effets paradoxaux et désinsatiables

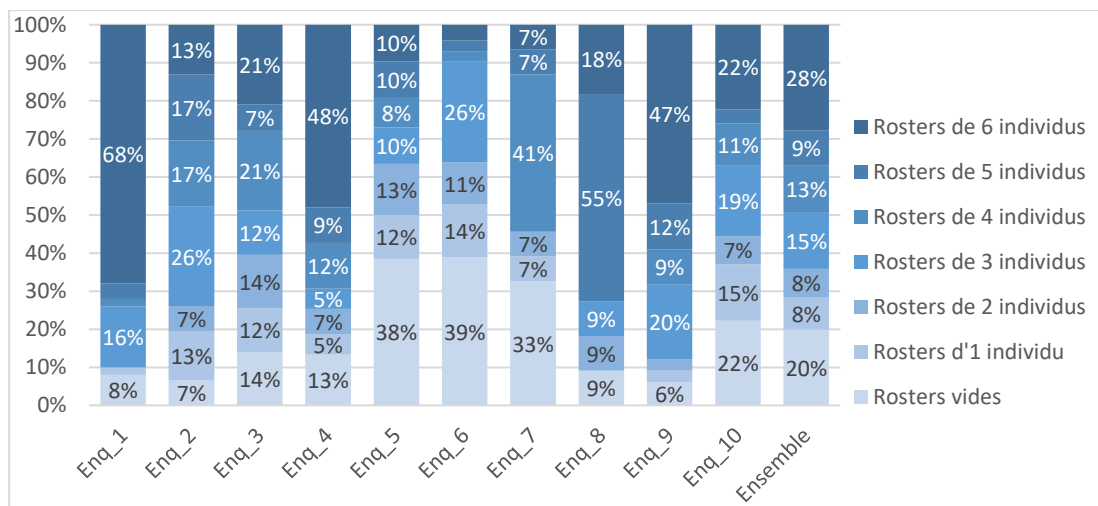
Première découverte : **la préférence pour les petits rosters**. Tandis que nos consignes aux enquêteur·rices insistaient sur l'importance de collecter de grands rosters (« collectez le maximum de citations dans vos rosters »), nous avons peu à peu réalisé que l'algorithme favorisait à l'inverse les petits rosters lors des échantillonnages. 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'aversion pour les doublons**. L'algorithme défavorise de surcroît les rosters contenant des doublons (un doublon est un individu déjà cité au préalable par un·e autre enquêté·e et identifié comme tel à l'issue de notre processus de matching). 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 II) 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. Le travail de collecte de ces informations identifiantes, pourtant difficiles à négocier, n'est donc pas valorisé.

Conséquence : des stratégies et des « biais enquêteurs » non négligeables

Ces tendances de l'algorithme sont contre-productives car les enquêteur·rices finissent 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 maximisant ses chances de voir ses rosters sélectionnés au tirage suivant (voir enq_6 ci-dessous). 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 enquêté·es disposaient sans doute de réseaux plus étendus que ce dont leur roster témoigne en définitive, et ses enquêté·e-s au roster vide avaient probablement un réseau, plus difficile à négocier, mais pas inexistant.

Distribution des rosters collectés en fonction de leur taille (hors graines) par enquêteur·rice



⇒ L'algorithme en mode *Search* considère que les répondants avec un grand roster et un roster avec doublon sont peu susceptibles d'être des nœuds-ponts, ils sont alors rarement prioritaires lors de l'échantillonnage alors qu'ils sont les plus coûteux à négocier et qu'ils sont nécessaires à la dynamique de progression dans le réseau. En conséquence, les stratégies pour minimiser ses efforts et maximiser ses chances de voir ses rosters échantillonnés induisent des biais de collecte importants.

Un algorithme de tirage difficile à s'approprier pour les enquêteur·rices

Le fonctionnement de l'algorithme suscite de nombreuses questions de la part des enquêteur·rices, tandis que l'équipe gestionnaire de collecte a eu également besoin de s'approprier les logiques sous-jacentes aux résultats des premiers tirages.

Ils ne comprennent pas (et à juste titre) comment « le hasard » peut à ce point faire focaliser le tirage sur quelques rosters, favorisant ainsi les enquêteurs qui les ont collectés.

Dès le second point mensuel, nous décidons donc (Géraldine Charrance et Aurélie Santos) de présenter un PPT pour leur expliquer le fonctionnement de l'algorithme (voir annexe 8). De plus, la modification de l'algorithme en décembre 2020 a permis de relativiser cette focalisation observée en début de collecte (voir partie VI.2.a).

En dépit de cette présentation et nombreux rappels de fonctionnement, des incompréhensions persistent. Une référente (point du 20/04/2021) ne comprend pas pourquoi, si l'algorithme cherche les personnes « rares » dans le réseau, il ne nous amène pas davantage vers des populations rares (Wenzhou) ? Il faut expliquer régulièrement que l'algorithme est « aveugle » à l'origine régionale ou tout autre caractéristiques sociodémographiques : c'est une rareté approchée par « le nombre de citation ». De plus, il y a eu des tirages de ce profil, mais comme ils tendent à davantage refuser, cela explique leur faible présence dans notre échantillon (ce n'est donc pas un effet du tirage).

c- Et complexe à faire comprendre aux enquêté·es

En janvier, les enquêteurs se sont plaints du **retour des vieux rosters**. Lors du point mensuel (26/01/2021), ils se sont consultés mutuellement pour voir comment les uns et les autres s'en sortaient pour convaincre des cités de participer alors que leur citant est difficile à re-mobiliser puisqu'il a été enquêté il y a plusieurs mois. Les enquêtés s'étonnent : « c'est encore en cours ? mais combien de temps cela va encore durer ? ». Ils ont la flemme de relancer leurs cités. Les enquêteurs tentent de s'expliquer comme ils peuvent : la collecte est ralentie à cause du Covid, on a du mal à recruter des

enquêtés à cause du Covid donc c'est important de nous aider... Une référente les prévient au moment du questionnaire, en expliquant que le tirage au sort peut se faire durant toute la collecte.

Ils font aussi part de la difficulté à **refuser les volontaires** qui ne sont pas tirés au sort. Les cités motivés pour participer ont tout le mal du monde à comprendre ces restrictions. Une enquêtrice donne un exemple (février 2021): « *Avant que j'ai fini de donner mes consignes, le citant a déjà envoyé les fiches de contact WeChat à ses amis proches et ils m'ont invitée. Donc trop tard on est lié, pas possible de refuser. Et un des cités m'a même appelé directement pour faire le questionnaire ! C'est vraiment dur de dire non* ».

De même une autre a été confrontée à une personne citée voulant être enquêtée de suite. Elle a donc dû commencer par refuser, tout rétro pédalant plusieurs semaines plus tard car cette personne a fini par être échantillonnée. Ce décalage dans la temporalité de la collecte, entre moment de la citation et moment d'un éventuel échantillonnage est très complexe à gérer sur le terrain.

3- Les difficultés liées au contexte de la pandémie

a- L'impact du COVID sur la collecte

Comme déjà mentionné dans diverses parties précédemment, nous avons connu un contexte très défavorable à l'enquête :

Notre population cible était d'autant plus difficile à joindre dans ce contexte de crise :

- Population très affectée par le COVID et peu disponible (inquiétude pour leur famille en Chine, difficultés liées à l'arrêt de certaines activités économiques, racisme anti-asiatique, etc.)
- Les enquêteur·rices ont observé une contraction du réseau de sociabilité des enquêté·es qui ne favorise pas les citations dans le roster
- Il y a des sous-groupes dont l'accès dépend d'un travail de terrain quasi-qualitatif (donc sur place en face-à-face): la communauté Wenzhou, les personnes en situation administrative irrégulière.

De plus, nos leviers de négociation ont largement été altérés par l'abandon du face-à-face :

- Plus grande difficulté à obtenir les infos de contact
- Pas de maîtrise de la négociation entre citant et cité (en face à face on peut espérer assister à ces négociations à la fin du questionnaire, et les orienter, mais c'est impossible en distanciel)
- Plus faible effet incitatif du chèque cadeau à distance car remis en différé.

b- Le passage au distanciel

Dès la fin septembre 2020, les enquêteur·rices et référentes font remonter que l'aggravation de la pandémie donne lieu à des annulations de RDV, de la méfiance et que certains s'auto-confinent. Le 28 octobre 2020 démarre le second confinement, nous obligeant à passer au mode distanciel.

Les modifications de protocole induites

- Envoyer des photos des cartes codes et le consentement avant la passation
- Le téléphone ne peut plus être utilisé en modem → il faut utiliser son wifi personnel
- Il n'est plus possible de scanner les QR code WeChat à distance → les enquêtés peuvent envoyer la "fiche contact" WeChat de leurs cités après la passation

- Il faut systématiquement passer par Isabelle Attané pour l'envoi postal des chèques cadeaux et des consentements papier. Il faut supprimer les mails comprenant des adresses postales régulièrement.
- Vers avril 2021, les enquêté·es viennent parfois chercher leur ticket cadeau à l'Ined, ce qui leur permet également de voir et vérifier ce qu'est l'Ined.

Toutefois, deux enquêtrices mentionnent un avantage du distanciel : cela permet de minimiser la perte de temps dans les transports (d'autant plus couteuse lorsque l'on fait face à un lapin) et les enquêté·es pouvaient également trouver cela plus facile pour s'organiser (certains ont maintenu ce choix même après la possibilité de retourner sur le terrain).

4- Les évolutions des consignes au fil du terrain

a- Les 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.

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 d'un montant de 15 euros pour le questionnaire, et de 20 euros pour le remplissage d'un roster dans lequel 6 contacts de l'enquêté·e sont renseignés. Il nous a semblé plus réaliste de distribuer le chèque roster à partir de 3 contacts cités.

b- La gestion des rosters incomplets

Les enquêteurs remontent des difficultés et hésitations sur la gestion des rosters incomplets. Combien de temps attendre dans l'espoir de récupérer des coordonnées de contact des personnes citées par un enquêté ? Nous n'avions pas prévu de protocole précis à ce sujet. Après examen par Géraldine Charrance du temps moyen passé avant la validation d'un roster au bout d'un mois de collecte, il est décidé de fixer un délai de 2 semaines et une concertation avec Aurélie Santos en cas de doute à l'issue de ce délai (27/10/2020).

c- La gestion des refus

Les refus explicites sont faciles à gérer. Les refus déguisés en évitement sont plus difficiles à arbitrer. Fin octobre 2020, après examen des conversions en statut « refus » et « enquêté » (10 jours) on fixe à 2 semaines le délai de passage en refus. Il faut une forte suspicion qu'il y a de l'évitement et un minimum d'effort réalisé sur ces cas (c'est à dire un certain nombre d'évènements déjà saisis dans l'appli). À l'issue des 2 semaines, faire un aux gestionnaires de collecte pour trancher ensemble le passage en refus (si tous les gens en attente basculent en refus au bout de 2 semaines, sans concertation, cela risque d'augmenter de beaucoup le taux de refus).

d- La gestion du suivi dans l'application

Nous réalisons en début de collecte qu'il est très compliqué de savoir si les rosters collectés sont complets ou en attente de nouvelles citations après relance au citant. Or l'envoi d'un roster à l'échantillonnage ne peut se faire que lorsque le roster est entier ou que l'enquêteur est sûr d'avoir collecté le maximum de lignes et que le citant ne donnera pas plus d'information. Pour résoudre ce problème nous avons ajouté une consigne : ajouter ligne vide pour signifier que le rosters est

incomplet. L'avantage est que l'on évite l'erreur (très pénible) d'envoyer à l'échantillonnage un roster qui ne doit pas l'être. L'inconvénient est que si l'enquêté ne donne finalement aucune autre ligne de roster, il faut demander de supprimer cette ligne « vide », en général la veille des échantillonnages.

Fin novembre lors du point mensuel (27/11/2020) nous rappelons aux enquêteur·rices de bien noter tous les événements de contact dans le listing « à enquêter » car nous constatons que les événements de type « relance auprès du citant » sont peu mentionnés, ce qui donne l'impression d'un faible effort sur les rosters incomplets, à tort.

e- Le suivi de l'affectation des échantillonnés

En cas de rupture de filiation, l'enquêteur A qui a enquêté le citant et l'enquêteur B en charge de l'exploitation de la fiche du cité peuvent être amenés à communiquer. En cours de collecte, nous constatons donc qu'il peut être intéressant que l'enquêteur A note des informations sur chaque cité concernant le degré de négociation atteint (pour l'enquêteur B en cas de transfert, ou pour lui-même pour mémoire) :

- Le cité est au courant de l'enquête et il est OK
- il est au courant mais réticent ou refuse
- il n'est pas au courant et le citant est OK pour qu'il soit contacté directement
- il n'est pas au courant et il faut d'abord revenir vers le citant avant de contacter le cité

Certaines particularités sur un cité ou un citant peuvent être mentionnées : par exemple la nature du lien entre citant et cité, ou si le citant ne veut pas qu'on mentionne son nom.

5- Les problèmes techniques

En décembre 2020 nous découvrons un nouvel obstacle sur WeChat : la barre de recherche de l'appli ne reconnaît pas les ID qui commencent par « wxid_ » alors qu'ils existent réellement. Nous avons d'abord cru à des faux WeChat donnés intentionnellement par les enquêté·es pour se débarrasser de l'enquête mais il s'avère que c'est un problème lié à l'application. De plus, nous découvrons que certains choisissent l'option de se masquer complètement sur WeChat et ne peuvent ainsi plus être l'objet d'une recherche.

En mai, l'application de collecte a connu de forts ralentissements en raison d'un problème technique sur le réseau de l'Ined. Nous avons dû ré-éditer des questionnaires papier pour les enquêteur·rices et prévoir un dédommagement pour le temps passé à la saisie ultérieure des questionnaires papier dans l'application de collecte (une fois celle-ci remise en état de marche).

6- Conclusion

La méthode NSM est une méthode innovante qui propose des avancées par rapport aux méthodes par chaînage plus classique de type RDS, sur le plan théorique. Le passage à la mise en pratique reste à ce jour très complexe. C'est une méthode peu utilisée et non stabilisée et qui ne dispose donc pas de guide pratique établi pour sa mise en œuvre. La transposition de l'algorithme initial en un programme exécutable en production a nécessité un travail considérable. En dépit de cette automatisation conséquente des processus de collecte (algorithmes de matching et de tirage, application web de collecte et de suivi conçue ad-hoc), la mise en place de cette méthode sur le terrain a nécessité de nombreux ajustements en début et en cours de collecte (algorithme, consignes, ajout de graines complémentaires) sans recours possible à des guidelines méthodologiques claires.

L'avancée sur le terrain repose sur la collecte des rosters et la négociation auprès des cités échantillonnées afin qu'ils acceptent l'enquête à leur tour. Ces étapes nécessitent une implication considérable des enquêteurs sur le terrain, d'autant plus auprès des profils les plus précaires. La dimension aléatoire et la temporalité des tirages produisent des effets peu intuitifs et difficiles à faire comprendre aux enquêtés.

De surcroît, l'algorithme produit des effets contre-productifs rendant l'expérience de terrain et la gestion de la collecte d'autant plus ardues. Les effets pervers de l'échantillonnage induisent des comportements d'enquêteur-rices qui, par stratégie ou découragement, peuvent aller à l'encontre de l'objectif initial de collecter des rosters de grande taille et d'identifier correctement les doublons. 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 sur lesquels repose la méthode NSM. Enfin, la méthode, à la frontière entre quanti et quali, nécessite impérativement un cadre de passation et de négociation en face-à-face, ce dont nous a privé la pandémie de Covid.

VII- Retours sur le questionnaire

1- Les questions difficiles à faire comprendre

Évaluation de son statut social : cette question est jugée difficile par la quasi-totalité des enquêteur-rices. Les enquêtés demandent des précisions, une définition plus claire de « statut social ». Cette dimension ne semble pas faire sens pour eux, ils peinent à l'évaluer. Ils pensent que cela renvoie uniquement au salaire, ou alors qu'ils n'ont pas de statut car ils sont étrangers, ou ils pensent « statut d'étudiant étranger ». Ou encore : certains savent avoir un statut « mauvais » par rapport aux français, mais jugent avoir un bon statut par rapport à d'autres chinois et ne savent donc que conclure. Dans un registre proche, certains ont des difficultés à objectiver et situer leur niveau de maîtrise du français.

Évaluation des relations avec d'autres groupes / communautés :

- En général, les enquêtés se plaignent souvent qu'il n'existe pas d'option « neutre » entre « plutôt bon » et « plutôt mauvais », entre « respectueux » et « indifférence », etc.
- Évaluation des relations entre les Chinois-es et les Français au quotidien : les enquêtés hésitent souvent entre « bienveillantes » et « indifférentes ».
- Les relations avec différentes populations chinoises (Wenzhou, Qingtian...) : les enquêtés ont tendance à penser aux individus qu'ils connaissent, mais non aux communautés dans leur ensemble.
- De plus, beaucoup de gens ne connaissent pas Qingtian et ils distinguent pas du tout les Chinois de Wenzhou et de Qingtian. Pour eux, c'est la même chose.

Évaluation de la taille du réseau : M_RESEAUCHIN : Combien de personnes chinoises ou d'origine chinoise fréquentez-vous à Paris ou en région parisienne ? Plusieurs enquêtrices font état de difficultés avec la notion de « fréquenter ».

2- Les questions sensibles

Les revenus : Quatre enquêteur-rices ont relevé des réticences liées aux questions sur les revenus actuels en France, et sur les biens détenus en Chine ou les transferts d'argent vers la Chine. Les

personnes travaillant au noir souhaitent que l'on coche « aucun revenus » pour coller à leur situation officielle, bien qu'ils perçoivent des revenus par ailleurs.

Trajectoire migratoire : Pour deux enquêteur·rices la question sur les raisons de l'émigration hors de Chine a pu susciter des réticences. Une autre précise qu'en cas d'arrivée en situation irrégulière sur le territoire français il peut être délicat de parler de sa trajectoire (pays où l'enquêté a vécu avant la France, lieu de résidence actuel car différent de son adresse officielle, etc.).

Profession des parents : Deux enquêtrices mentionnent la question de la profession des parents aux 15 ans d'ego (E_PROFPER, E_PROFMER) comme étant sensible. Même en ayant expliqué que cela servirait à connaître la catégorie socio-professionnelle, une donnée cruciale pour l'analyse sociologique, certains enquêtés donnaient des réponses vagues ou des refus.

Les discriminations : deux enquêtrices pensent que les discriminations sont sous déclarées car les enquêté·es semblent avoir honte de les avouer. Les enquêtés racontent des situations désagréables ou considérées par eux-mêmes comme des discriminations parfois, puis ils en minimisent les conséquences, ils produisent des discours qui justifient ces comportements et demandent de ne pas les noter (ex : il est normal d'être traité ainsi car on vit en étranger chez d'autres ; je suis tolérant je pardonne ; mon niveau de français est insuffisant c'est normal que l'on perde patience avec moi, etc.). Par ailleurs, une enquêtrice pense qu'il y a un enjeu spécifique de la part des hommes qui ne souhaitent pas « perdre la face » (expression qui revient souvent pour décrire les réactions des enquêté·es chinois·es) et qui pousse à la minimisation des difficultés rencontrées au regard de leur insertion dans la société française.

Situation conjugale : une enquêtrice relève le fait de devoir répondre aux questions sur la situation conjugale dès le début du questionnaire semble intrusif pour certains. Il a fallu leur expliquer que leur réponse servirait à faire une analyse sociologique, par exemple, à analyser la relation entre la situation conjugale et la volonté de s'installer en France.

Santé psycho : une enquêtrice relate que la question sur la santé psychologique n'est pas forcément sensible mais peut être gênante. Certains enquêtés coupent court : « je n'ai pas de problème sur ce point ».

Appartenance à des organisations ou parti politique : une enquêtrice relève l'aspect potentiellement sensible de l'engagement politique.

VIII- Liste des annexes

Annexe 1 : Réseaux des graines de l'enquête ChIPRe

Annexe 2 : Schéma des résultats du test d'octobre 2019

Annexe 3 : Délibération de la Cnil

Annexe 4 : Plan des journées de formation

Annexe 5 : Supports de communication

Annexe 6 : Analyses statistiques menées sur la base des rosters

Annexe 7 : Évolutions de la méthode d'échantillonnage

Annexe 8 : Fonctionnement de l'algorithme de matching –présentation pour les enquêteur·rices