



De la théorie à la pratique : Quelques enseignements de la mise en œuvre de la méthode « Network sampling with memory » pour enquêter auprès des immigrants chinois en Ile-de-France

Aurélie Santos, Géraldine Charrance, Marine Bondon, Paul Cochet & Thomas Merly-Alpa
Ined, service des enquêtes et sondages

Plan de la communication

- **NSM, une méthode innovante et ambitieuse pour l'étude de populations cachées**
 - Notre population d'intérêt : Les immigrés chinois en Ile-de-France
 - Les méthodes d'enquêtes par réseau, particulièrement adaptées à ce type de populations
 - Les apports de la variante NSM
- **La complexe mise en œuvre de la méthode NSM sur le terrain**
 - Finaliser le protocole d'échantillonnage et de collecte sur le « tas »
 - Découvrir un algorithme aux effets contre-intuitifs et parfois désincitatifs
 - Expérimenter un contexte sanitaire très défavorable aux enquêtes en face-à-face
- **Traitements post-enquête et premiers résultats**

Plan de la communication

- **NSM, une méthode innovante et ambitieuse pour l'étude de populations cachées**
 - Notre population d'intérêt : Les immigrés chinois en Ile-de-France
 - Les méthodes d'enquêtes par réseau, particulièrement adaptées à ce type de populations
 - Les apports de la variante NSM
- **La complexe mise en œuvre de la méthode NSM sur le terrain**
 - Finaliser le protocole d'échantillonnage et de collecte sur le « tas »
 - Découvrir un algorithme aux effets contre-intuitifs et parfois désincitatifs
 - Expérimenter un contexte sanitaire très défavorable aux enquêtes en face-à-face
- **Traitements post-enquête et premiers résultats**

Les populations difficiles à enquêter



- Caractéristiques des populations difficiles à enquêter :
 - De petite taille
 - Absence de base de sondage (appartenance à la population non identifiable dans la base de sondage)
 - Appartenance des individus à la population pas toujours évidente à identifier (population cachée voire stigmatisée)
- Nécessité de recourir à des méthodes moins conventionnelles
- choisir une méthode d'enquête adaptée aux spécificités de la population d'intérêt

Les populations difficiles à enquêter



- Enquêtes en population générale pour capter ces populations peu adaptées car :
 - Trop peu de personnes de la population d'intérêt pour mener des analyses
 - Risque de sous-déclaration
- **Les immigrés chinois en Ile-de-France**
 - 69 répondants nés en Chine ou HK pour l'enquête en population générale TeO
 - Recensement ou EDP comportent peu d'informations pour caractériser finement les populations immigrées
 - Suspicion forte de défaut de couverture (migrants en situation irrégulière par exemple)

Les méthodes d'enquêtes par réseau



- Enquêter au sein de réseau en mettant à contribution les répondants pour en recruter de nouveaux
- Pourquoi recourir à ce type de méthodes ?
 - Ne nécessite pas de base de sondage
 - Répondants plus susceptibles de participer et de répondre avec honnêteté lorsqu'ils sont invités à participer par des pairs
 - Volonté de comprendre la structure et l'organisation de la population d'intérêt

Mais, ces méthodes conduisent à des estimations biaisées.

Les méthodes d'enquêtes par réseau



Cependant, la probabilité d'être échantillonné peut être estimée par le nombre de liens de chaque répondant avec les autres membres de la population sous certaines hypothèses:

- Population formée d'un seul réseau (tous les individus sont connectés, théorie de la chaîne de Markov)
- Liens entre individus sont réciproques
- Individus recrutés aléatoirement parmi les connaissances

Variance d'échantillonnage dépend de la taille de l'échantillon et de la structure du réseau (effet de cluster).

Network sampling with memory : Améliorer l'échantillonnage à partir d'un réseau en recueillant des données sur le réseau. Au fur et à mesure de l'enquête, la liste des personnes désignées tend à ressembler à la liste complète des membres de la population.

Network sampling with memory (NSM)



Principes généraux

- Point de départ : une enquête exploratoire pour trouver des graines aux profils très diversifiés
- Un questionnaire + un roster (grille de contacts, jusqu'à 6 personnes) pour chaque enquêté (récompensé par incitation financière)
- Un algorithme de tirage sélectionne un échantillon d'individus au sein du réseau dévoilé (Algorithme NSM)
- Les enquêteur·rice·s doivent enquêter uniquement ces individus échantillonnés
- Chaque semaine le tirage est reconduit sur la base de sondage constituée grâce aux rosters collectés

→ Le tirage au sein du réseau dévoilé permet en théorie de lisser les biais habituellement reprochés aux méthodes de collecte en réseau de type RDS

Network sampling with memory (NSM)



Search Mode (*Mode par défaut*)

Network sampling with memory (NSM)



Search Mode (*Mode par défaut*)

« Explorer le réseau »

Diriger l'enquête vers des parties du réseau encore inexplorées.

- On calcule la probabilité d'être un nœud-pont de chaque répondant (basée sur la proportion de personnes citées une seule fois parmi ses « contacts »).
- On sélectionne un individu parmi les 5 individus les plus susceptibles d'être des nœuds-ponts
- On tire aléatoirement un de ses amis parmi ceux cités une seule fois et non enquêtés.

Network sampling with memory (NSM)



*Examen du niveau
d'exploration du réseau*

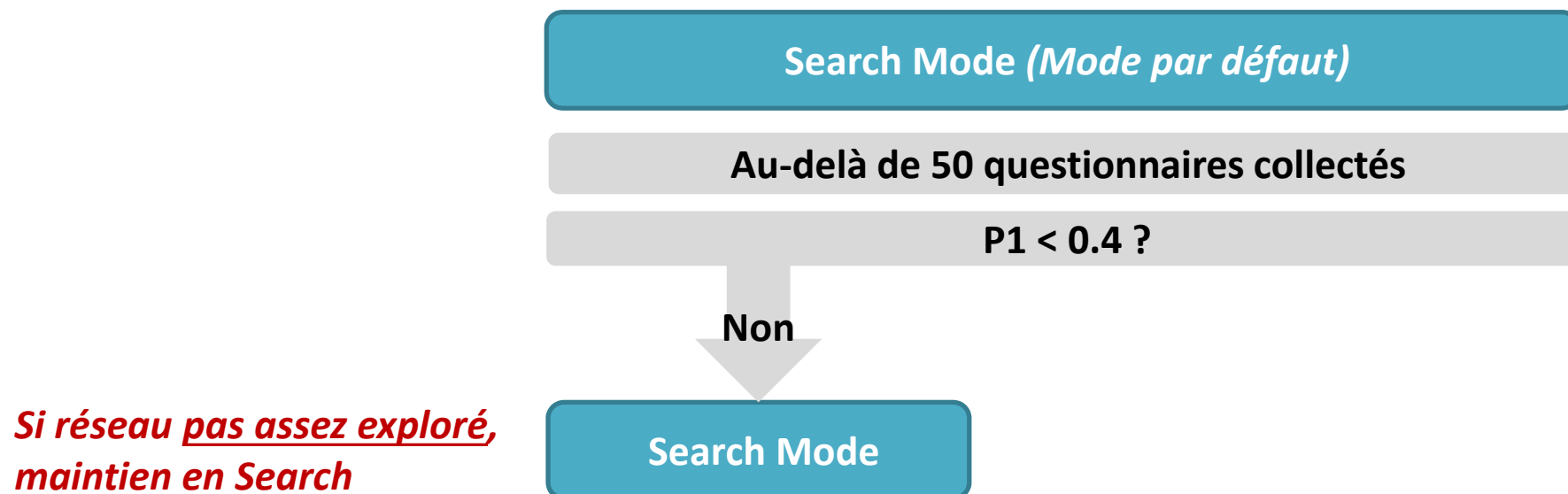
Search Mode (*Mode par défaut*)

Au-delà de 50 questionnaires collectés

$P1 < 0.4$?

P1 = part de personnes citées une seule fois et non enquêtées parmi toutes les personnes dévoilées

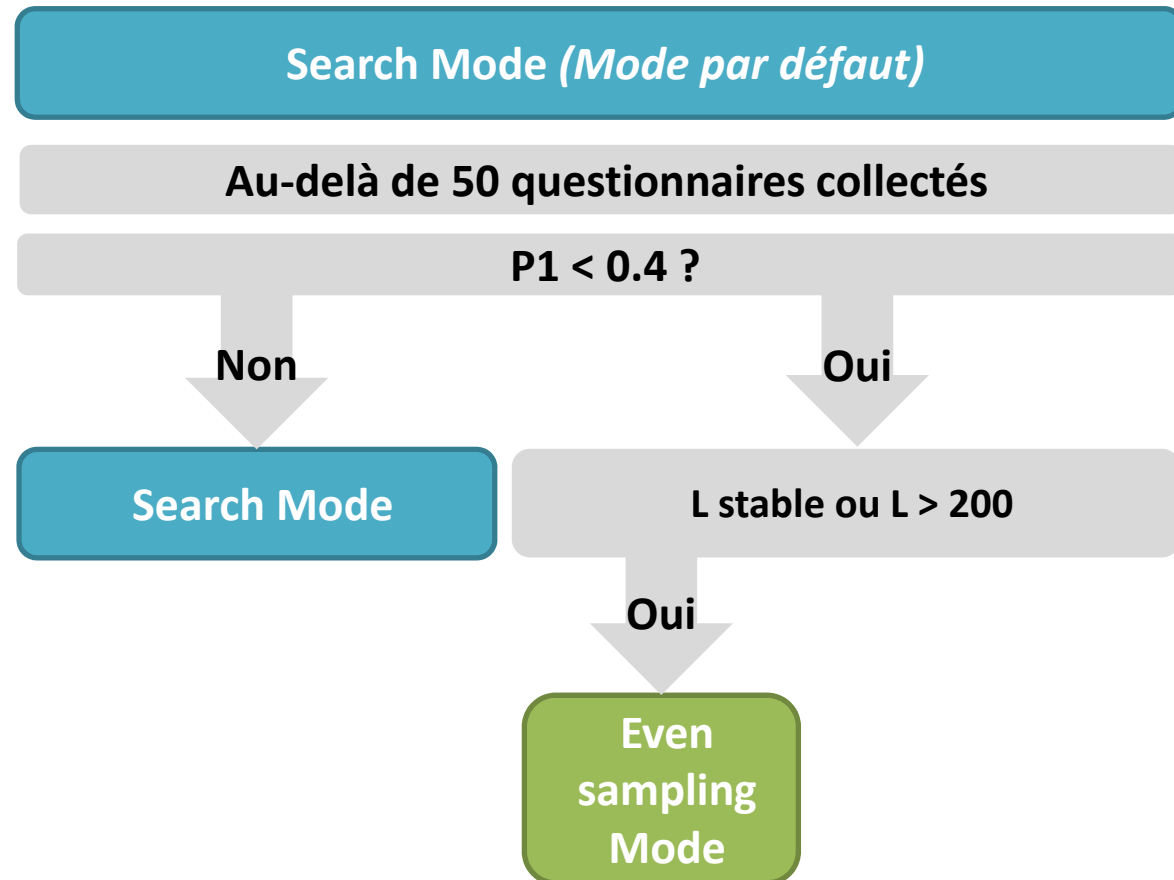
Network sampling with memory (NSM)



Network sampling with memory (NSM)



*Si réseau assez exploré
et de taille suffisante ou
stable,
homogénéisation de
l'exposition au tirage*



Network sampling with memory (NSM)

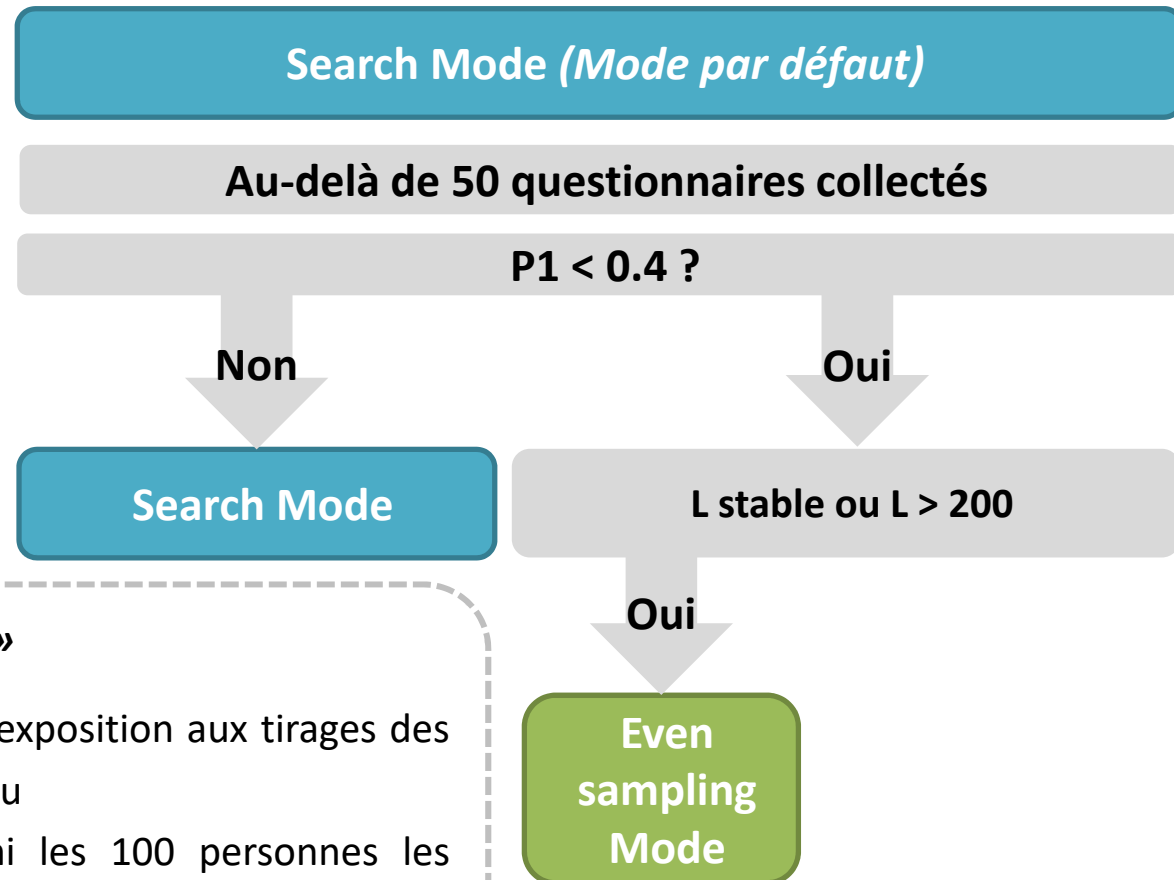


*Si réseau assez exploré
et de taille suffisante ou
stable,
homogénéisation de*

« équilibrer l'échantillonnage »

L'objectif est d'homogénéiser l'exposition aux tirages des individus présents dans le réseau

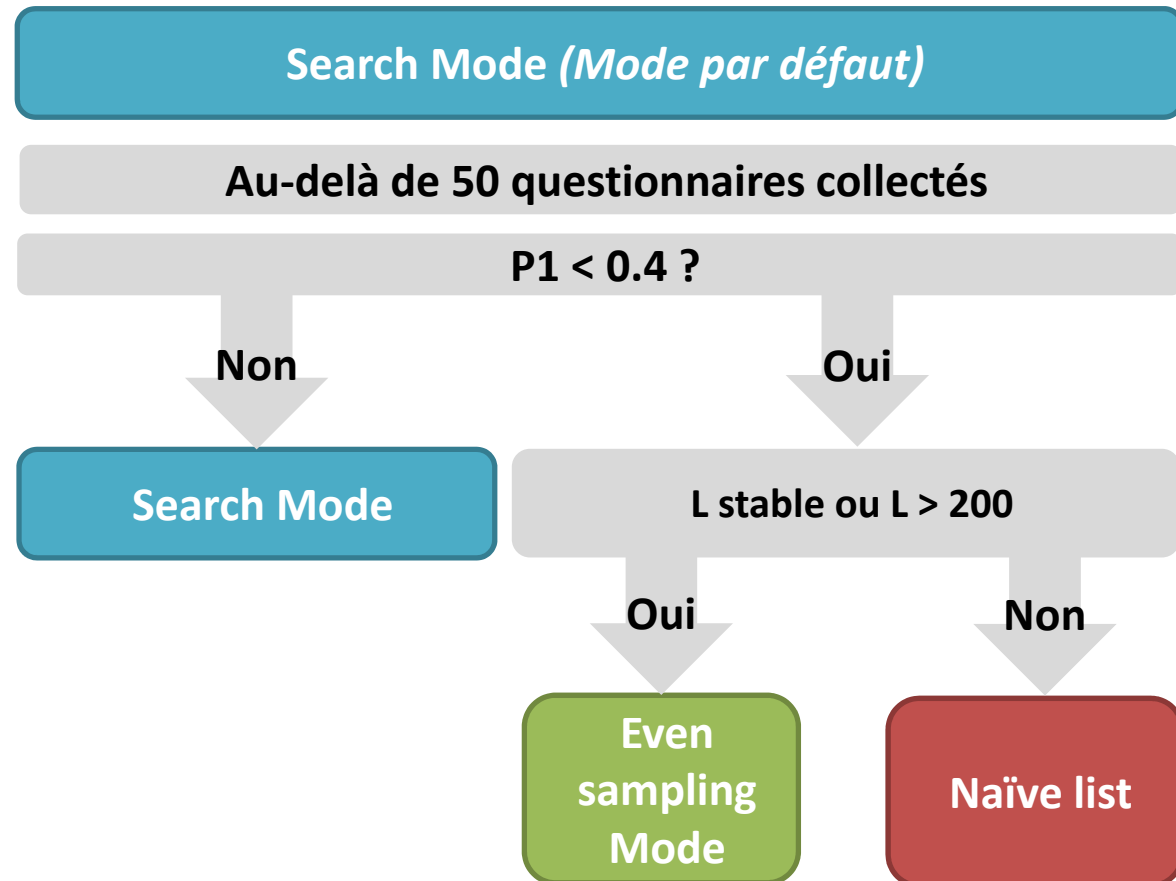
- Tirage aléatoire simple parmi les 100 personnes les moins exposées jusqu'à alors



Network sampling with memory (NSM)



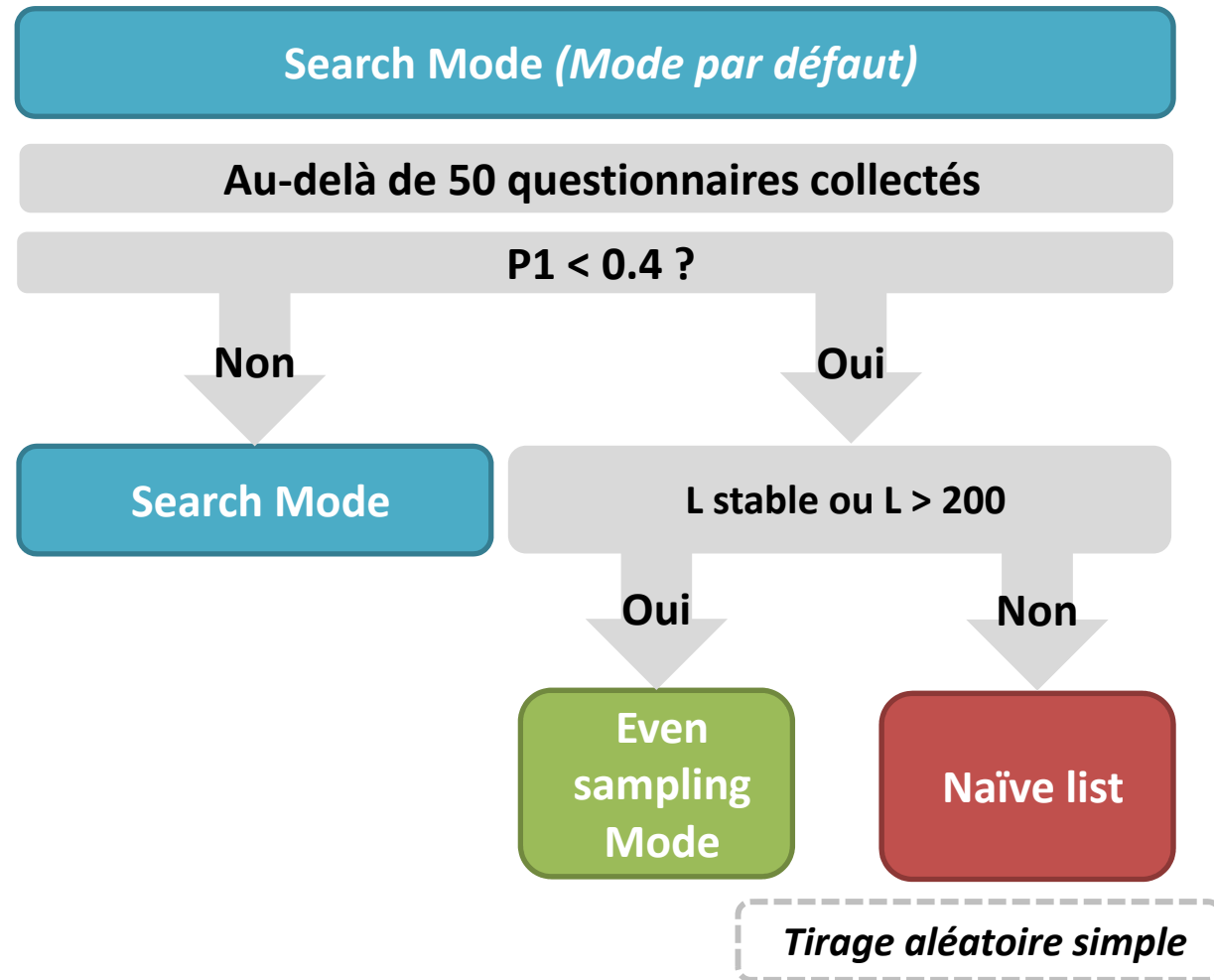
*Si réseau assez exploré
et taille inférieure à 200
et non stable*



Network sampling with memory (NSM)



*Si réseau assez exploré
et taille inférieure à 200
et non stable*



Plan de la communication

- **NSM, une méthode innovante et ambitieuse pour l'étude de populations cachées**
 - Notre population d'intérêt : Les immigrés chinois en Ile-de-France
 - Les méthodes d'enquêtes par réseau, particulièrement adaptées à ce type de populations
 - Les apports de la variante NSM
- **La complexe mise en œuvre de la méthode NSM sur le terrain**
 - Finaliser le protocole d'échantillonnage et de collecte sur le « tas »
 - Découvrir un algorithme aux effets contre-intuitifs et parfois désincitatifs
 - Expérimenter un contexte sanitaire très défavorable aux enquêtes en face-à-face
- **Traitements post-enquête et premiers résultats**

- **Finaliser le protocole d'échantillonnage et de collecte sur le « tas »**
 - Déterminer le dimensionnement des tirages:
 - À quelle fréquence réaliser des échantillonnages?
 - De quelle taille?
 - Comme les répartir entre enquêteur·rice·s?

 - Modifier l'algorithme de tirage pour mieux cibler les éligibles au mode search

 - Modifier le protocole face aux réalités de terrain:
 - Extension de la durée de collecte
 - Assouplissement des critères pour les incentives
 - Ajout de graines

- **Découvrir les effets contre-intuitifs et désincitatifs de l'algorithme**

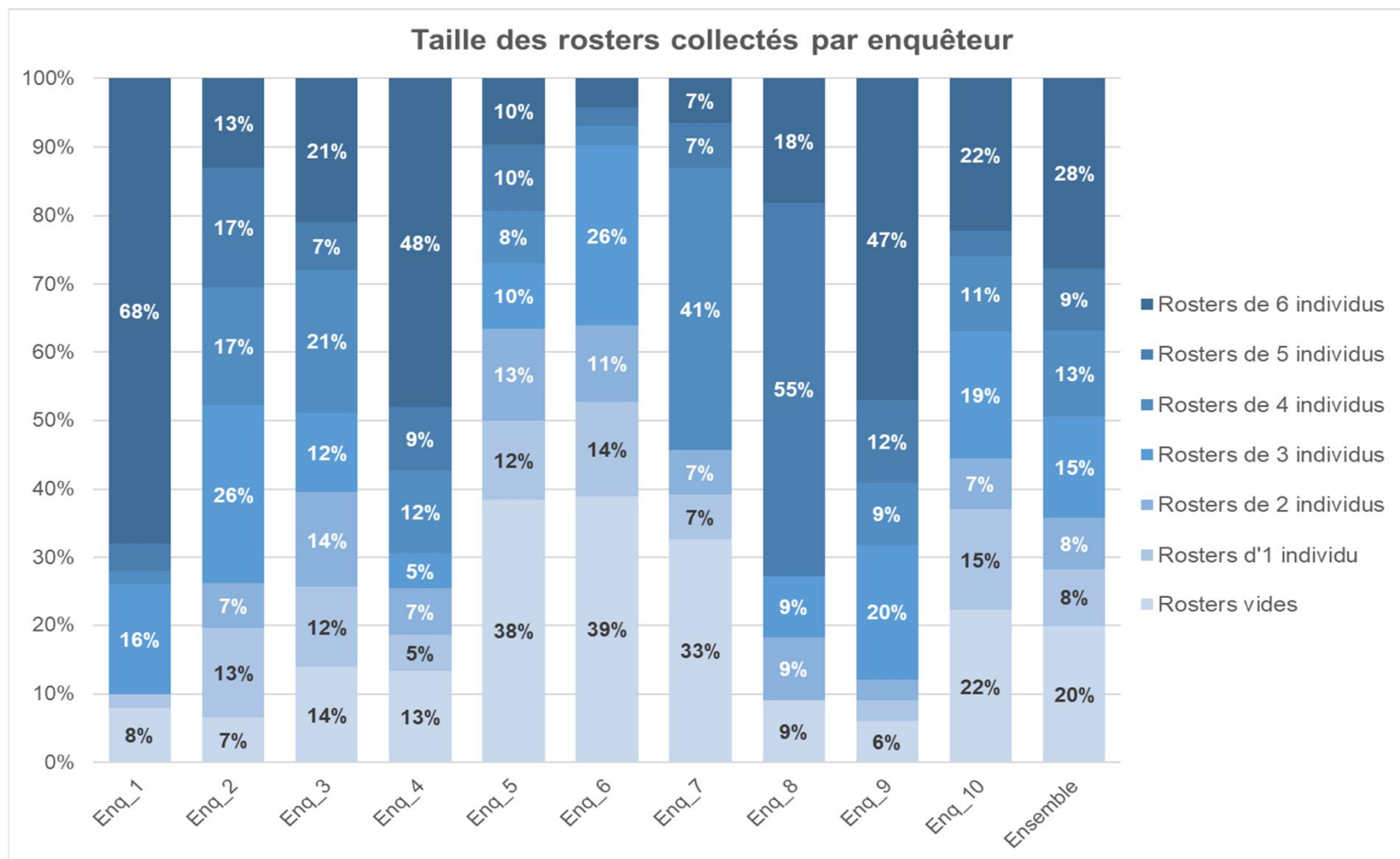
- **La préférence pour les petits rosters** alors que les consignes insistent sur la nécessité d'un effort collectif pour dévoiler le maximum de liens du réseau étudié

- **L'aversion pour les doublons** alors que les consignes insistent sur leur identification

→ 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

- Les « biais enquêteur·rice·s » induits (*cf. graph suivant*)

Les biais enquêteur·rice·s en conséquence



- **Expérimenter un contexte sanitaire très défavorable aux enquêtes en face-à-face**
 - Une population d'autant plus difficile à joindre dans ce contexte de crise
 - Population très affectée par le COVID et peu disponible
 - Contraction du réseau de sociabilité
 - 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

 - Des leviers de négociation largement altérés par l'abandon du face-à-face
 - Grande difficulté à obtenir les infos de contact
 - Pas de maîtrise de la négociation entre citant et cité
 - Faible effet incitatif du chèque cadeau à distance

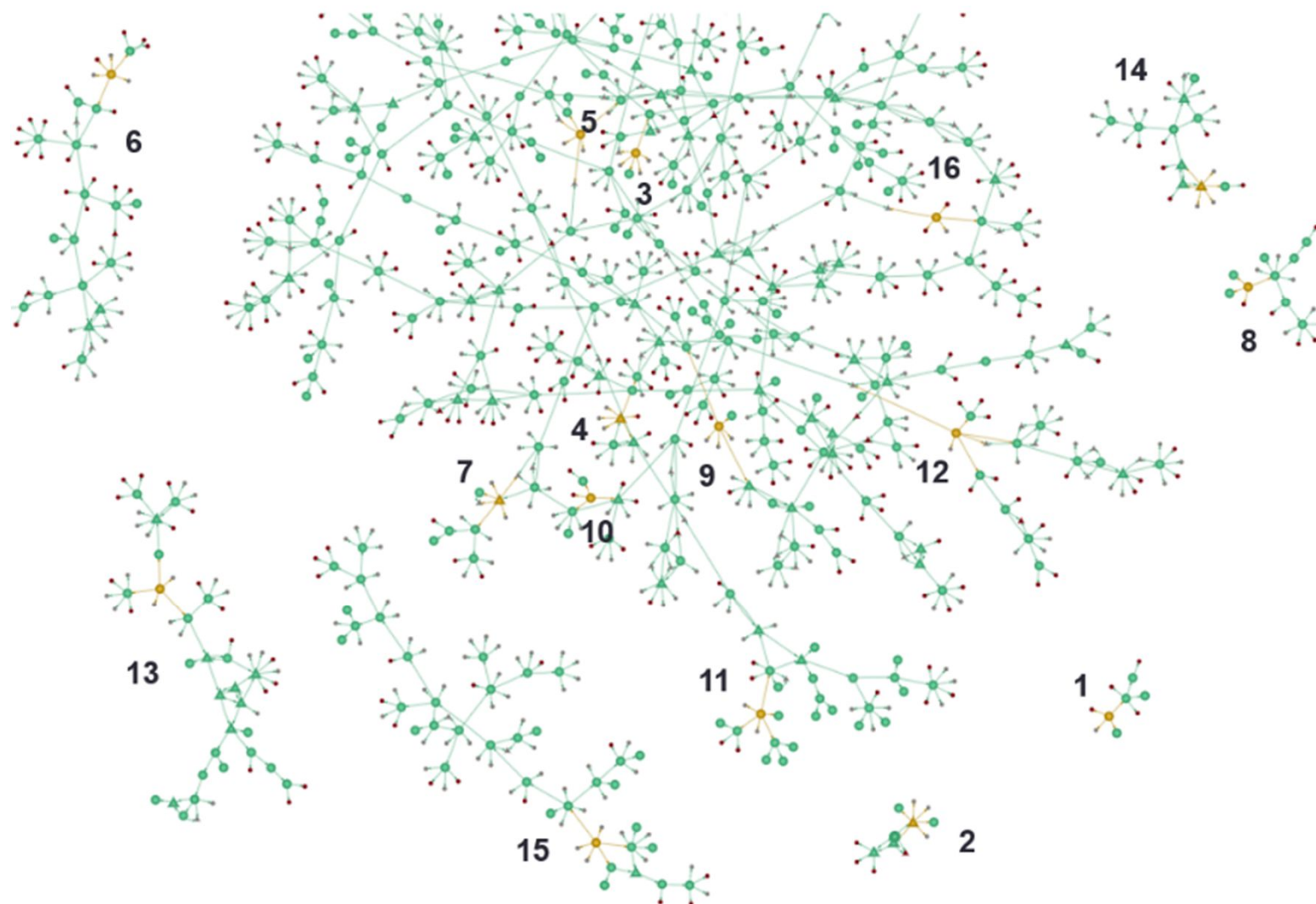
Plan de la communication

- **NSM, une méthode innovante et ambitieuse pour l'étude de populations cachées**
 - Notre population d'intérêt : Les immigrés chinois en Ile-de-France
 - Les méthodes d'enquêtes par réseau, particulièrement adaptées à ce type de populations
 - Les apports de la variante NSM
- **La complexe mise en œuvre de la méthode NSM sur le terrain**
 - Finaliser le protocole d'échantillonnage et de collecte sur le « tas »
 - Découvrir un algorithme aux effets contre-intuitifs et parfois désincitatifs
 - Expérimenter un contexte sanitaire très défavorable aux enquêtes en face-à-face
- **Traitements post-enquête et premiers résultats**

Bilan de la collecte

- Nombre de questionnaires: **501**
- Nombre de citations dans le réseau: **1689**
- Nombre d'individu unique dans le réseau: **1523**
- Taille moyenne des rosters: **3,37 citations**
- Durée moyenne pour obtenir un roster complet non vide: **4,7 jours** (max 6 jours)
- Taux de réponse: **60% des échantillonnés** (non hors-champ)
- Indicateur P1: **0,6** (seuil de sortie du Search = 0,4)
- Nombre de tirages: **45**

Bilan de la collecte



Post-traitements

- Quelle formule de poids utiliser ?
 - Le Bootstrap utilisé ex-post par les concepteurs? → hypothèses sous-jacentes incertaines
 - La formule de poids pensée ex-ante par les concepteurs? → forte dispersion des poids
- Quelle stratégie pour le calage?
 - Abandon de la correction de la non-réponse par GRH
 - Calage sur le RP2018
- Beaucoup d'incertitudes !
 - Quel impact de la structure des graines?
 - Quel impact du maintien en Search sans passage par le mode Even sampling?

Conclusion

- Contexte de crise sanitaire a eu un impact sur :
 - L'acceptation de l'enquête
 - La recommandation par les pairs
 - La portée des incitations financières
 - Le rythme de collecte
 - Le travail des enquêteurs
- Cependant, certaines difficultés sont intrinsèques à la méthode NSM :
 - Imprévisibilité de la charge enquêteur
 - Préférence pour les petits rosters & aversion pour les doublons
- NSM apparait comme une méthode conçue à partir de réseaux théoriques ou déjà dévoilés qui peine à mesurer le coût pratique du processus même de dévoilement du réseau.
- Méthode semble prometteuse mais plus adaptée à une population plus facile à enquêter et de plus petite taille.