

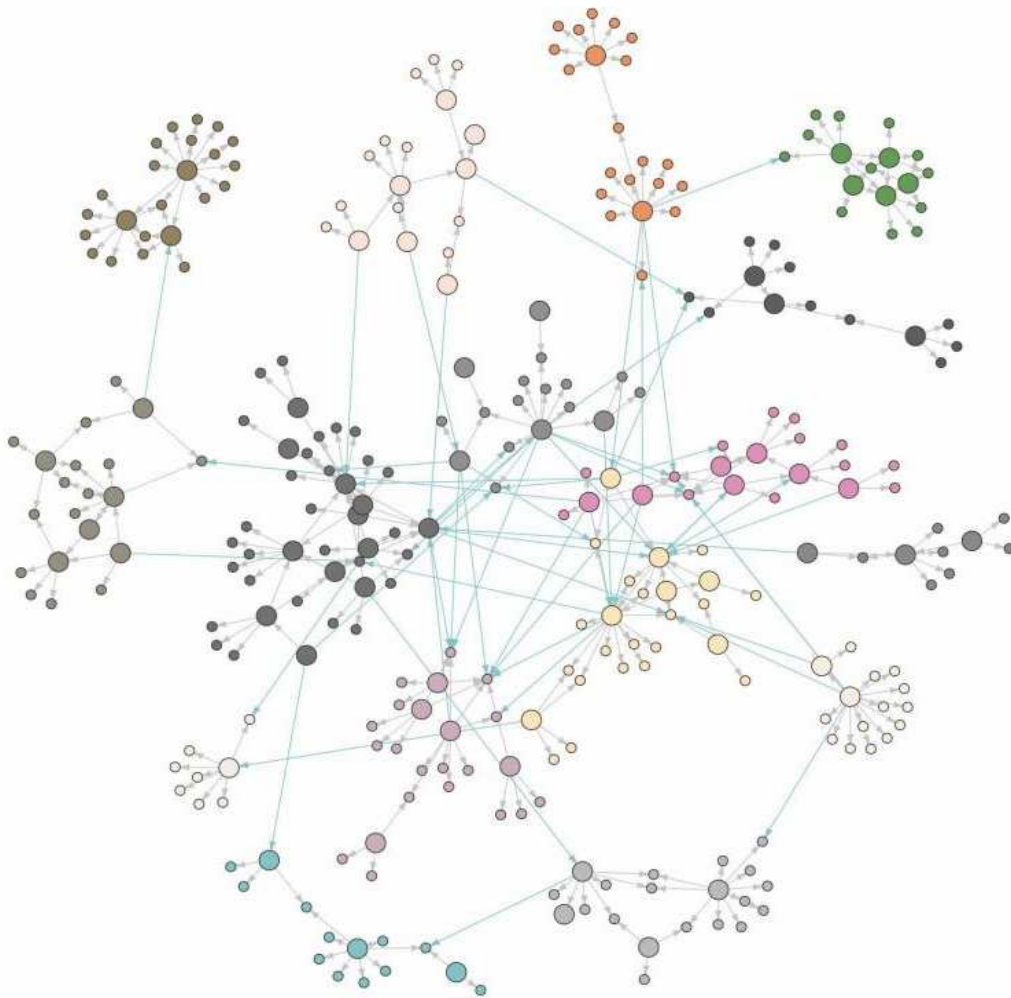
Mardi 11 juin 2019, Après-midi

COLLECTE ET GESTION DES DONNÉES RELATIONNELLES

Vanesse Labeyrie et Sarah Ouadah

Formation « Introduction à l'Analyse de Réseaux »

Projet CoEx, 11-15 juin 2019, Dakar



Objectifs du cours

- Connaître les différentes étapes à suivre pour mettre en œuvre une étude sur les réseaux sociaux.
- Etre capable de choisir la stratégie d'échantillonnage adapté à votre question, et connaître ses limites.
- Pouvoir anticiper les différentes difficultés que vous pouvez être amenés à rencontrer lors de la collecte de données.
- Connaître comment organiser vos données dans le tableur, et comment créer un objet réseau avec le package R Igraph.

Différentes étapes

1. Avant de collecter les données

- a. définir la stratégie d'échantillonnage
- b. construire le questionnaire
- c. considérations éthiques

2. Pendant la collecte de données: s'adapter aux imprévus

3. Après la collecte de données

- a. Rentrer les données dans un tableur
- b. Créer et manipuler l'objet réseau avec le package R igraph

1a.Définir la stratégie d'échantillonnage

Avant de collecter les données

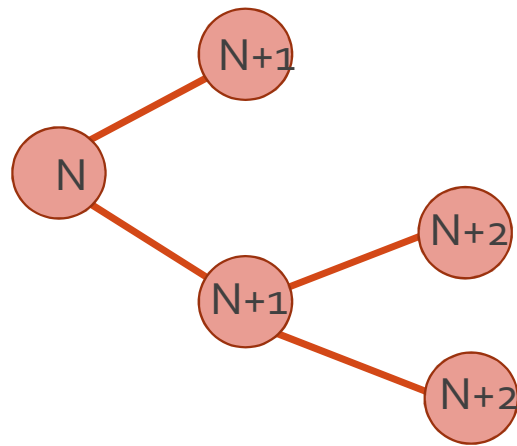
Différentes stratégies de collecte de données

Selon l'approche choisie (positionnelle VS structurale), cela va changer considérablement la stratégie d'échantillonnage.

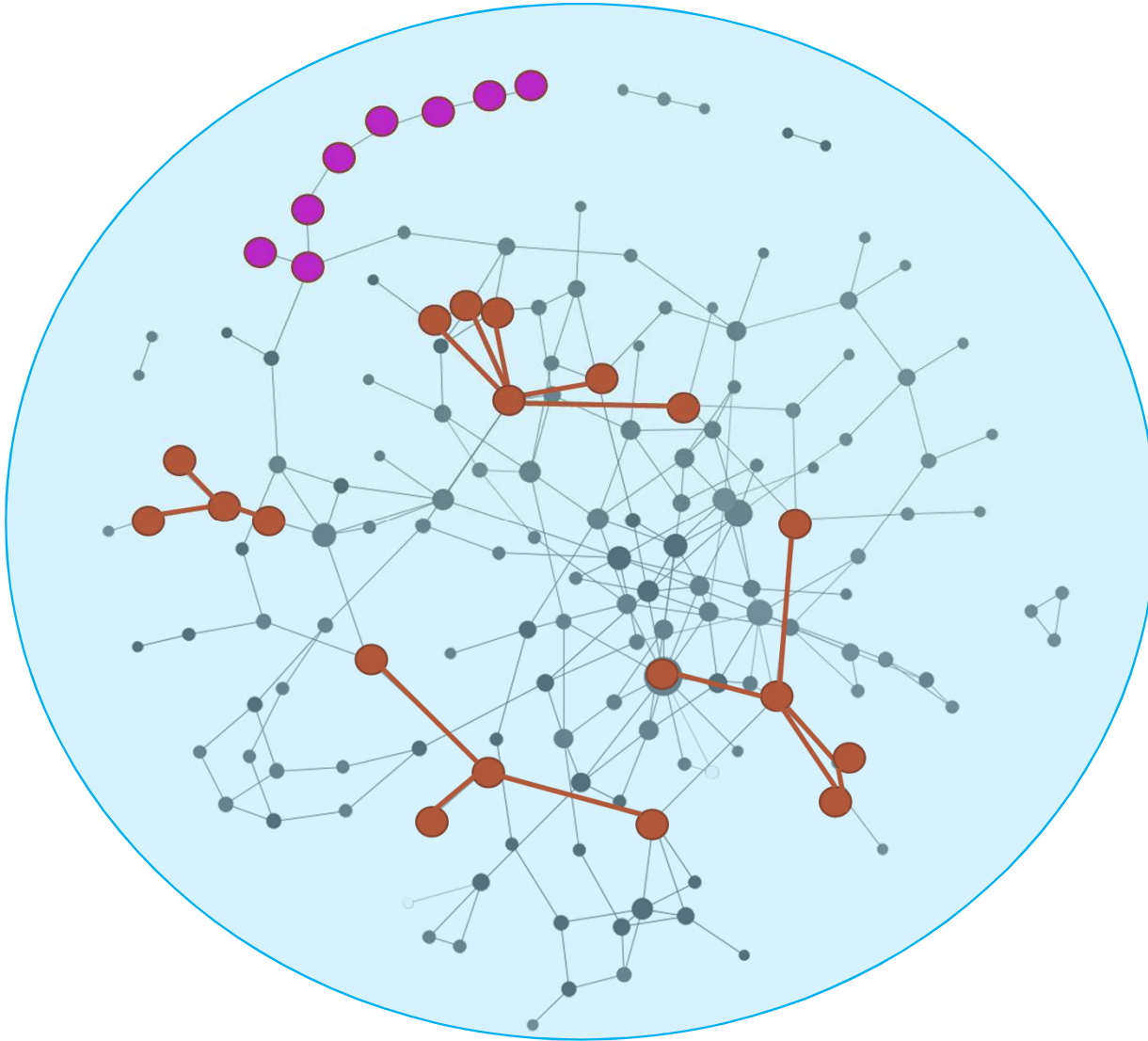
- Pour une **approche individuelle**: il est possible de choisir un échantillon de nœuds dans une population de grande taille suivant les méthodes classiques (aléatoire, stratifié).
- Pour une **approche structurale**: il est nécessaire d'avoir un échantillonnage exhaustif à l'intérieur des limites que l'on a fixé. Ce type d'enquête ne peut donc être conduit que sur des groupes de taille assez limitée.

Différentes stratégies de collecte de données

Lorsque l'on cherche à étudier un **processus de diffusion** partant d'un nœud connu, il est possible de documenter le réseau avec une stratégie d'échantillonnage de proche en proche, de type « boule de neige ».



Différentes stratégies de collecte de données



Approche individuelle:
ego network

Diffusion: boule de neige

Réseau complet

Différentes stratégies de collecte de données

Si un réseau est trop « creux », c'est-à-dire qu'il présente une faible densité de liens, les approches structurales ne peuvent pas être mises en œuvre.

Il peut donc être utile de réaliser des enquêtes préliminaires pour estimer la densité du réseaux a priori.

Différentes stratégies de collecte de données

- **Générateur de nom:**

« de qui avez-vous reçu des semences l'année dernière ? »

- **Roster (liste pré-définie):**

« Parmi les personnes de cette liste, indiquez quelles sont celles qui vous ont donné des semences l'année dernière, et quelles sont celles qui ne vous en ont pas donné »

→ Cela suppose de disposer d'un recensement exhaustif de tous les « nœuds » dans les limites que l'on a fixé.

1b. Construire le questionnaire

Avant de collecter les données

Formuler les questions

- Eviter les questions vagues, décrire et **préciser** la relation étudiée le mieux possible, et ancrer les questions sur des objets et des événements concrets:

Par exemple, plutôt que de demander « de qui avez-vous reçu des semences ? », demander « je vois que vous cultivez telle variété, auprès de qui avez-vous obtenu les semences ? »

- Tenir compte de la **mémoire** des personnes enquêtées:

Par exemple, il est plus simple de se souvenir des variétés reçues que données.

- Tenir compte des **normes sociales** locales:

Par exemple, selon les cultures, il est honteux de dire que l'on a reçu ou au contraire que l'on a donné.

Enquêtes: Questionnaires

1. Renseigner les caractéristiques de EGO et lui attribuer un identifiant unique.
 2. Lister chaque lien établi par EGO et lui attribuer un identifiant unique.
 3. Pour chaque lien, renseigner l'identité de ALTER afin de pouvoir l'identifier sans ambiguïté et faire correspondre le bon identifiant.
 4. Décrire les attributs de ALTER (si on ne l'enquête pas).
 5. Pour chaque lien, décrire la relation entre ALTER et EGO en renseignant les attributs pertinents au regard de la question.
- ➔ Etablir une typologie des attributs des nœuds et des liens au préalable, à partir des pré-enquêtes.

Questionnaires

ID, Nom et attributs de **EGO**

Qualifier le lien

Qualifier l'ALTER

ID Lien	direction	Variété	Quantité	Année	Nature	ID ALTER	Nom ALTER	Relation	Genre	Village
1	Donné	Mugana	1	1998	Don	1	Joseph M.	Parent	M	Kabururu
2	Reçu	Mugeta	0,5	2000	Vente	2	John T.	Ami	M	Gatanga
3	Reçu	Kaguru	2	2010	Vente	3	Theresa C.	Voisin	F	Kabururu
4	Donné	Gadam	10	2017	Troc	4	Robert F.	Aucune	M	Kabururu

Qu'est ce qui circule

Quand

Comment

Enquêtes: le qualitatif

Ne pas oublier le qualitatif !!

Fondamental pour préciser les hypothèses à tester, et interpréter les résultats des analyses de réseau.

Quelles sont les normes, règles, habitudes, formelles ou non, qui influencent les interactions étudiées ?

1c. Questions éthiques

Avant de collecter les données

Quelques principes éthiques à respecter

- S'assurer que l'étude conduite ne risque pas de porter préjudice aux individus ou au collectif étudié.
- Obtenir l'accord des autorités locales, administratives, et coutumières.
- Obtenir l'accord des participants : consentement libre et éclairé.
- Anonymiser les réponses, ne publier aucune des informations permettant d'identifier les individus ou de localiser les ménages, voir les villages précisément.
- L'enquêteur doit strictement respecter la confidentialité concernant les données non-anonymisées.

2. Gérer les imprévus

Pendant la collecte des données

Des « imprévus » classiques

- Impossibilité d'enquêter certains individus dans le cadre d'une approche par réseau complet.
- Refus de donner des informations concernant certains types de relations.

Et vous, quelles difficultés avez-vous rencontré ?

3a. Rentrer les données sur tableur

Après avoir collecté les données

Entrée des données dans le tableur

Fichier des nœuds comprenant:

- un identifiant unique pour chaque nœud.
- les attributs de chaque nœud

Fichier des liens comprenant:

- un identifiant unique pour chaque lien
- le nœud « source » (EGO) et le nœud « cible » (ALTER)
- les attributs de chaque lien

Préparer les données dans le tableur: étape 1

Etape de construction des fichiers de donnée dans Excel

- Créer un fichier des liens

ID Lien	ID EGO	direction	Variété	Quantité	Année	Nature	ID ALTER	Relation
1	l1	Donné	Mugana	1	1998	Don	l2	Parent
2	l1	Reçu	Mugeta	0,5	2000	Vente	l3	Ami
3	l1	Reçu	Kaguru	2	2010	Vente	l4	Voisin
4	l1	Donné	Gadam	10	2017	Troc	l5	Aucune
5	l2	Reçu	Mugana	1	1998	Don	l1	Parent
6	l2	Donné	Mugeta	2	2000	Don	l5	Ami
7	l2	Donné	Mugana	1	2000	Don	l5	Ami

Préparer les données dans le tableur: étape 2

Etape de construction des fichiers de donnée dans Excel

- Créer un fichier des nœuds

ID individu	Nom	Genre	Village	Statut enquête
I1	Albert C.	M	Kabururu	EGO
I2	Joseph M.	M	Kabururu	EGO
I3	John T.	M	Gatanga	ALTER
I4	Theresa C.	F	Kabururu	ALTER
I5	Robert F.	M	Kabururu	ALTER

Préparer les données dans le tableur: étape 3

Etape de « Nettoyage » dans le fichier Excel (selon les données):

- Enlever les citations réciproques (si on a documenté les relations bidirectionnelles)

ID Lien	ID EGO	direction	Variété	Quantité	Année	Nature	ID ALTER	Relation	
1	l1	Donné	Mugana	1	1998	Don	l2	Parent	
2	l1	Reçu	Mugeta	0,5	2000	Vente	l3	Ami	
3	l1	Reçu	Kaguru	2	2010	Vente	l4	Voisin	
4	l1	Donné	Gadam	10	2017	Troc	l5	Aucune	
5	l2	Reçu	Mugana	1	1998	Don	l1	Parent	
6	l2	Donné	Mugeta	2	2000	Don	l5	Ami	
7	l2	Donné	Mugana	1	2000	Don	l5	Ami	

Il est judicieux de calculer le niveau de « consensus », i.e les relations citées par les deux acteurs :
« A à dit avoir reçu de B, et B à dit avoir donné à A »

Préparer les données dans le tableur: étape 4

Etape de « Nettoyage » dans le fichier Excel (selon les données):

- Tri selon la direction du lien

ID Lien	ID EGO	direction	Variété	Quantité	Année	Nature	ID ALTER	Relation
2	l1	Reçu	Mugeta	0,5	2000	Vente	l3	Ami
3	l1	Reçu	Kaguru	2	2010	Vente	l4	Voisin
5	l2	Reçu	Mugana	1	1998	Don	l1	Parent

1	l1	Donné	Mugana	1	1998	Don	l2	Parent
4	l1	Donné	Gadam	10	2017	Troc	l5	Aucune
6	l2	Donné	Mugeta	2	2000	Don	l5	Ami
7	l2	Donné	Mugana	1	2000	Don	l5	Ami

Préparer les données dans le tableur

Etape de « Nettoyage » dans le fichier Excel (selon les données):

- Aboutir à un fichier où toutes les relations on la même direction: Donneur - Receveur

ID Lien	ID Donneur	ID receveur	Variété	Quantité	Année	Nature	Relation
1	l1	l2	Mugana	1	1998	Don	Parent
2	l1	l3	Mugeta	0,5	2000	Vente	Ami
3	l1	l4	Kaguru	2	2010	Vente	Voisin
4	l1	l5	Gadam	10	2017	Troc	Aucune
5	l2	l1	Mugana	1	1998	Don	Parent
6	l2	l5	Mugeta	2	2000	Don	Ami
7	l2	l5	Mugana	1	2000	Don	Ami

Les étapes 3 et 4 peuvent être sautées si on a documenté les relations dans une seule direction (par exemple seulement « donné » ou seulement « reçu »)

Préparer les données dans le tableur: étape 5

Etape de « Nettoyage » dans le fichier Excel (selon les données):

- Enlever les liens multiples si l'analyse ne porte pas sur cela.

ID Lien	ID EGO	direction	Variété	Quantité	Année	Nature	ID ALTER	Relation
1	l1	Donné	Mugana	1	1998	Don	l2	Parent
2	l1	Reçu	Mugeta	0,5	2000	Vente	l3	Ami
3	l1	Reçu	Kaguru	2	2010	Vente	l4	Voisin
4	l1	Donné	Gadam	10	2017	Troc	l5	Aucune
5	l2	Reçu	Mugana	1	1998	Don	l1	Parent
6	l2	Donné	Mugeta	2	2000	Don	l5	Ami
7	l2	Donné	Mugana	1	2000	Don	l5	Ami

Mise en application (Vanesse)

A vos fichiers des données !

3b. Créer et manipuler un objet réseau dans R

Après avoir collecté les données

Créer et manipuler un objet réseau avec igraph (R)

Le package R Igraph permet de créer un objet « réseau » en important les fichiers de données issus du tableur en format .csv

On peut ensuite manipuler cet objet, et effectuer des analyses .

Mise en application *(Sarah)*

- Initiation à R / rappel des bases.
- Créer et manipuler un objet réseau avec le package R Igraph.