

Les enjeux de la mise en forme des données de réseaux personnels ?

Wilda JEAN BAPTISTE¹

Résumé

Nous avons tous déjà entendu cet adage : « *Nous sommes la somme des cinq personnes que nous fréquentons le plus* », qui illustre l'analyse des réseaux personnels, une étude centrée sur un individu et son entourage [1]. Un réseau personnel est composé d'une personne que l'on nomme « *ego* » et d'un ensemble de personnes (« *alter* ») avec lesquelles il interagit ainsi que les relations existant entre ces derniers [2].

Les données collectées sur ego et son réseau sont initialement enregistrées dans un *data frame* dont les lignes représentent un ego et les colonnes, un attribut (sexe, âge, relations (*alter*), interconnaissances entre alter, etc.). Cependant le format brut des données ne permet pas de réaliser une analyse à l'échelle d'ego et des alters. Pour y remédier, les fonctions *pivot_longer()* (ou *pivot_wider()*) du package *tidyr* sous le logiciel R permettent de restructurer les données en produisant une base centrée sur ego et une base centrée sur les alters et leurs interconnaissances. On passe alors d'un format long à un format large qui permet de réaliser de la visualisation de réseaux sous forme de graphes, de calculer des indicateurs clés comme la densité d'un réseau.

Mots-clés : Ingénierie – Data – Package – Reproductibilité

Développement

L'analyse de réseaux vise à étudier le réseau d'une personne (*ego*), composée de ses relations de différents types professionnelles, personnelles, etc. Les données de réseaux sont recueillies principalement à l'aide de deux modes de recueil : par entretiens semi-directifs (face à face) pour la collecte de données qualitatives et par des questionnaires en lignes auto-administrées. Pour identifier les relations significatifs, des générateurs de noms, qui sont des questions, sont utilisés. Par exemple :

De nombreuses personnes ressentent le besoin de parler et de discuter de choses importantes avec d'autres personnes. Si vous avez des problèmes personnels – par exemple, concernant une personne proche ou quelque chose qui vous importe – avec qui en parleriez-vous ?

Ego peut également renseigner des informations complémentaires sur ses alters (type de lien : familial, amical, lieu de rencontre, fréquence des interactions, etc.).

Ces données sont initialement organisées dans un *data frame* (tableau) où dans chaque ligne décrit des informations propres à ego et chaque colonnes, un attribut d'ego (sexe, âge, lieu de nais-

¹Laboratoire Interdisciplinaire Solidarités, Sociétés, Territoires (LISST), wilda.jean-baptiste@univ-tlse2.fr

sance, etc.) et des alters (lien avec ego, âge, sexe, etc.) ainsi que les données d'interconnaissances (« L'alter A connaît-il l'alter B ? »). Comme indiquée précédemment l'analyse du réseau personnel s'opère à l'échelle d'ego mais également d'alter. Par exemple, il est difficile d'extraire directement les relations entre *alters* ou d'étudier la densité du réseau sans restructurer les données.

Pour en permettre l'analyse, les données sont restructurées à l'aide des fonctions `pivot_longer()`, du package **tidyr**, pour passer d'un format *large* (centré sur *ego*) à un format *long* (centré sur les *alters* ou leurs attributs) :

```
pivot_longer(  
  data = data_frame,  
  cols = c(nom_alters_1, nom_alters_2), # Colonnes à pivoter  
  names_to = "code_relation", # Nouvelle colonne pour les noms des colonnes  
  d'origine  
  values_to = "nom_alter" # Nouvelle colonne pour les noms des alters  
)
```

Les données transformées permettent de visualiser les réseaux sous forme de graphes où les noeuds symbolisent ego et ses alters, et les liens, les relations entre eux [3] et de calculer des indicateurs clés comme la densité d'un réseau ou le nombre moyen de relations citées par ego, la centralité, etc.. La mise en forme des données soulève des défis comme la gestion des alters pouvant être cités à plusieurs reprises, des données d'interconnaissances entre alters.

Cette présentation sera l'occasion de partager les solutions mise en oeuvre, à l'aide de R pour traiter ces défis et pourquoi pas l'ouverture à une nouvelle approche.

Bibliographie

- [1] C. Bidart, « Réseaux personnels et processus de socialisation », *Idées économiques et sociales*, vol. 169, n° 3, p. 8-15, déc. 2012, doi: 10.3917/idee.169.0008.
- [2] G. Favre et M. Grossetti, « Les réseaux personnels en France ont-ils changé ? Une comparaison entre 2001 et 2017 », *Revue française de sociologie*, vol. 62, n° 2, p. 167-208, 2021, doi: 10.3917/rfs.622.0167.
- [3] L. Saglietto, « Quelques points de repères dans l'étude des réseaux par la théorie des graphes », *Netcom*, vol. 20, p. 195-216, janv. 2006, doi: 10.3406/netco.2006.1654.