

# Manipulation de données bibliométriques via {managHAL}

Martin AMIENS

2024-04-07

## Martin AMIENS

- Stagiaire au sein de l'unité MIA - Paris-Saclay.
- Etudiant en première année du master Bio-Informatique et Bio-Statistiques à Paris-Saclay.
- Tuteurs : Julie AUBERT, Pierre BARBILLON

## Plan

- Contexte du stage
- Objectifs du stage
- Accessibilité et généricité de {managHAL}
- introduction aux Réseaux
- Ajout de fonctionnalités à {managHAL}
- Perspectives et avenir du package

## Contexte

- DRIT : Direction des Recherches et de l'Innovation Technologique

. . .

- Sciences Ouvertes : sont “la diffusion sans entrave des publications et des données de la recherche.” (plan-national-pour-la-science-ouverte,2021)

. . .

- HAL : Hyper Articles en Ligne



Figure 1: Crédit : CCSDForge

...

- {ManagHAL} : Package R d'extraction de données bibliométriques de HAL

### Objectifs du stage

- Accessibilité et généricité de {managHAL}

...

- Ajout de fonctionnalités à {managHAL}

### Accessibilité et généricité de {managHAL}

**Exemple d'URL** : "https://api.archives-ouvertes.fr/search/hal/?q=structId\_i:1002311&fq=publicationDate\_01-01T00:00:00Z%20TO%202023-01-01T23:59:59Z]&fq=docType\_s:(COMM+OR+ART ...  
&sort=producedDate\_tdate+desc&fl=docid,halId\_s,version\_i,docType\_s,citationFull\_s ...  
&defType=edismax&rows=1000"

## Accessibilité et généricité de {managHAL}

### Tab A

```
#| code-fold: true
```

```
data_RH_csv_Example <- load_team_table(filter_id = TRUE, date_cols = c(6,7), filepath_or_url
```

Warning in load\_team\_table\_csv(filter\_id = filter\_id, date\_cols = date\_cols, :  
NAs introduits lors de la conversion automatique

```
# afin de ne garder que les lignes avec un idhal correct
data_RH_csv_Example <- data_RH_csv_Example[which(!is.na(data_RH_csv_Example$idhal)),]
data_RH_csv_Example <- as_tibble(data_RH_csv_Example)
data_RH_csv_Example
```

```
# A tibble: 3 x 13
```

	civilite	nom	prenom	statut	rattachement	debut_contrat	fin_contrat
	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
1	M.	ADJAKOSSA	Éric	CEC	AgroParisTech	""	""
2	Mme	AUBERT	Julie	IR	INRAE	""	""
3	M.	BARBILLON	Pierre	PR	AgroParisTech	""	""

```
# i 6 more variables: financement <chr>, equipe <chr>, unite <chr>,
#   orcid <chr>, idhal <dbl>, adresse_mail <chr>
```

### Tab B

```
#| code-fold: true
```

```
date_min = "01/01/2022"
date_max = "01/01/2023"
```

```
# Publications à partir d'un id de laboratoire.
```

```
HAL_publis_Labo <- managHAL::HAL_extract_csv(id = 1002311, date_min, date_max, type_id = "stru
HAL_publis_Labo <- as_tibble(HAL_publis_Labo)
HAL_publis_Labo[1,]
```

```
# A tibble: 1 x 11
```

docid	halId_s	version_i	docType_s	citationFull_s	citationRef_s
<int>	<chr>	<int>	<chr>	<chr>	<chr>

```

1 4440523 hal-04440523          1 COMM      Isabelle Lebert, Maxim~ Colloque fin~
# i 5 more variables: publicationDate_tdate <chr>,
#   authFullNamePersonIDIDHal_fs <chr>, structAcronym_s <chr>,
#   structId_i <chr>, structHasAlphaAuthIdHalPersonid_fs <chr>

```

```

# Publications à partir d'ids auteurs
HAL_publis_auteurs <- managHAL::HAL_extract_csv(id = data_RH_csv_Example[,12][2,] , date_min
HAL_publis_auteurs <- as_tibble(HAL_publis_auteurs)
HAL_publis_auteurs

```

```

# A tibble: 4 x 8
  docid halId_s      version_i docType_s citationFull_s      citationRef_s
  <int> <chr>          <int> <chr>      <chr>          <chr>
1 3781287 hal-03781287          1 COMM      Wenfan Cao, Maxime Lec~ FOODMICRO 20~
2 3694338 hal-03694338          1 COMM      Wenfan Cao, Maxime Lec~ 23ème éditio~
3 3811607 hal-03811607          1 COUV      Julie J. Aubert, Pierr~ Statistical ~
4 4019111 hal-03604836          1 ART       Sana Romdhane, Aymé Sp~ The Internat~
# i 2 more variables: publicationDate_tdate <chr>,
#   authFullNamePersonIDIDHal_fs <chr>

```

## Introduction aux Réseaux

Un réseau est composé de :

- noeuds (entités)
- arêtes (interactions entre ces entités)

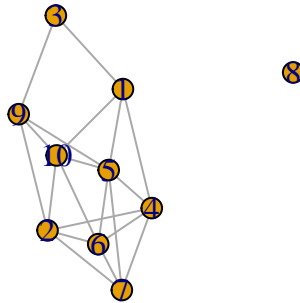
## Réseau Unipartite

```

# Création d'un réseau unipartite
set.seed(42) # Pour la reproductibilité
unipartite <- erdos.renyi.game(10, 0.3)
unipartite_graph <- plot(unipartite, main = "Réseau Unipartite")

```

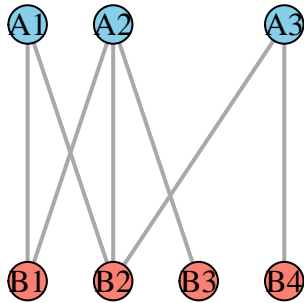
## Réseau Unipartite



## Réseau Bipartite

```
# Création d'un réseau bipartite Noms des nœuds
nodes_A <- c("A1", "A2", "A3")
nodes_B <- c("B1", "B2", "B3", "B4")
# Liste des arêtes
edges <- c("A1", "B1", "A1", "B2", "A2", "B1", "A2",
           "B2", "A2", "B3", "A3", "B2", "A3", "B4")
# Création du graphe
bipartite <- graph(edges, directed = FALSE)
V(bipartite)$type <- bipartite_mapping(bipartite)$type
plot(bipartite, layout = layout_as_bipartite, vertex.color = c("skyblue",
  "salmon")[V(bipartite)$type + 1], vertex.label.color = "black",
  vertex.shape = "circle", vertex.size = 30, edge.width = 2,
  main = "Réseau Bipartite")
```

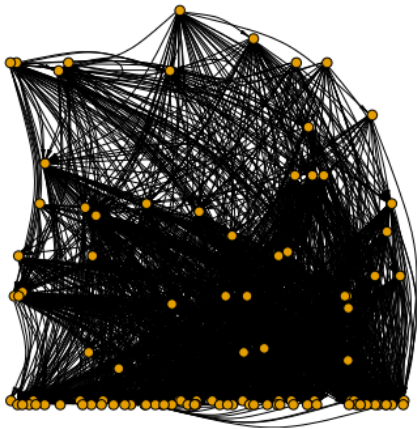
## Réseau Bipartite

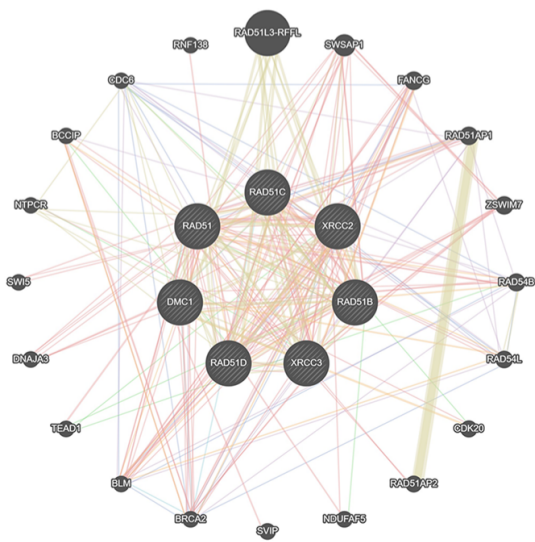


### Exemples

A : Exemple de Réseau trophique (réseau d'interactions des espèces chiliennes), Crédit : (Kéfi et al. 2016)

B : RAD51 et son réseau d'interactions de sa famille de gènes, Crédits : (Li et al., 2021)





## Ajout de fonctionnalités à {managHAL}

construction d'un réseau de co-auteur et SBM

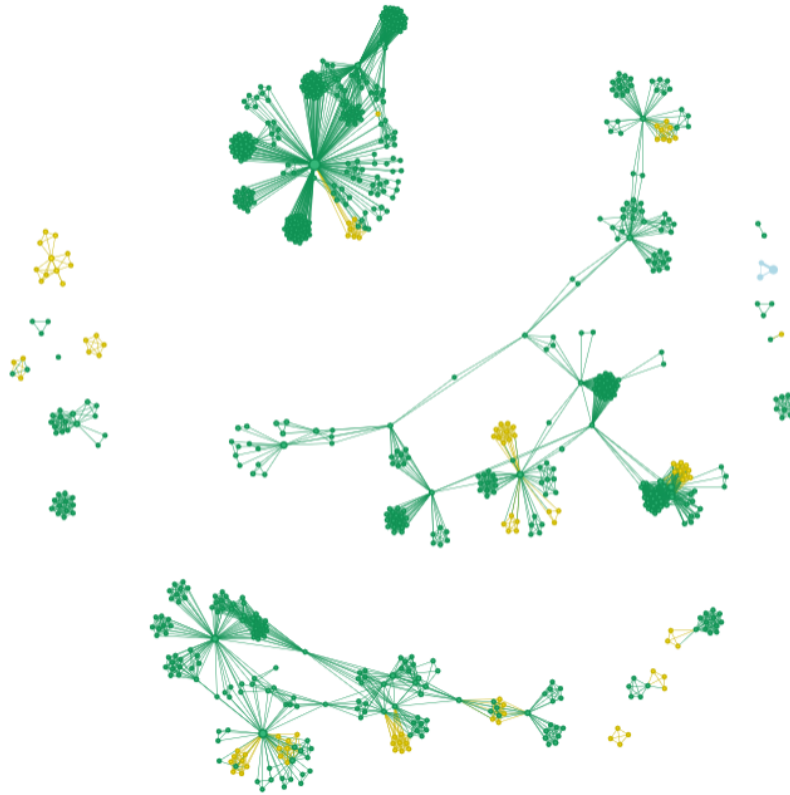


Figure 2: Réseau complet de co-auteurs de l'unité MIA - PS



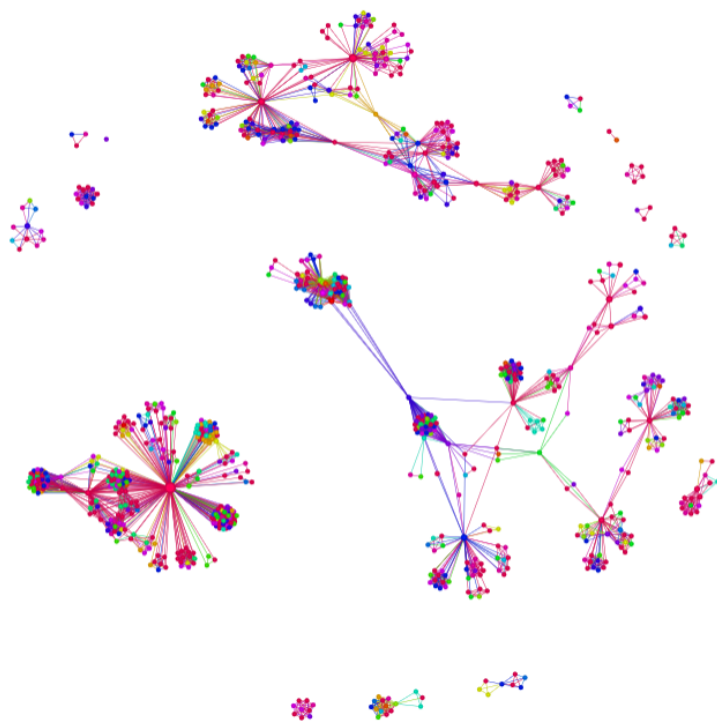


Figure 3: Réseau complet de co-auteurs de l'unité MIA - PS coloré par appartenance aux groupes

## Ajout de fonctionnalité à {managHAL}

### Création de sous-groupe

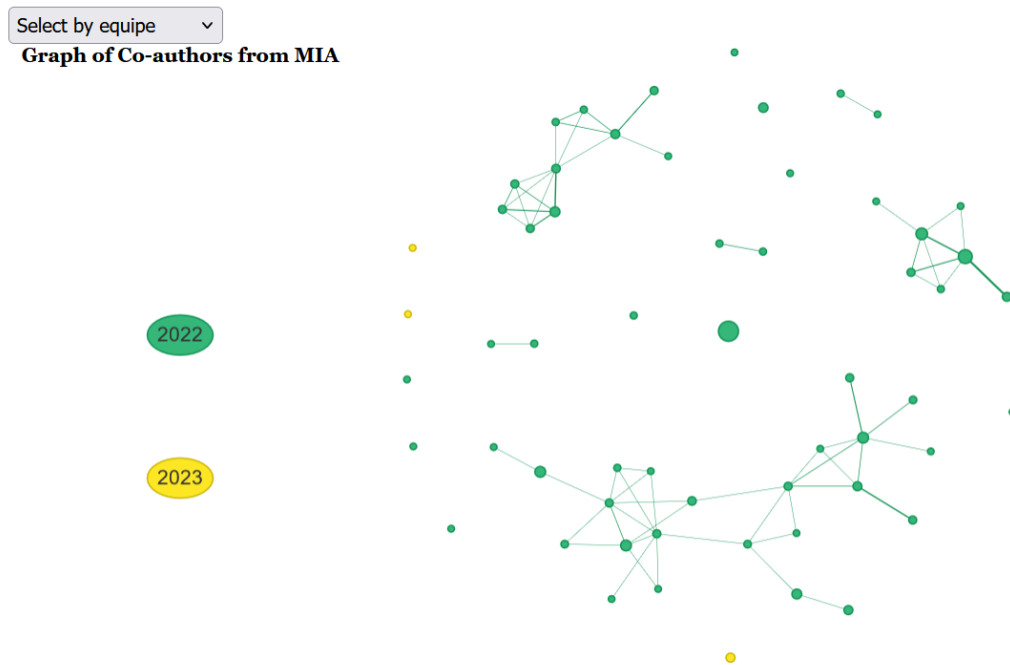


Figure 4: Réseau partiel de co-auteurs de l'unité MIA - PS

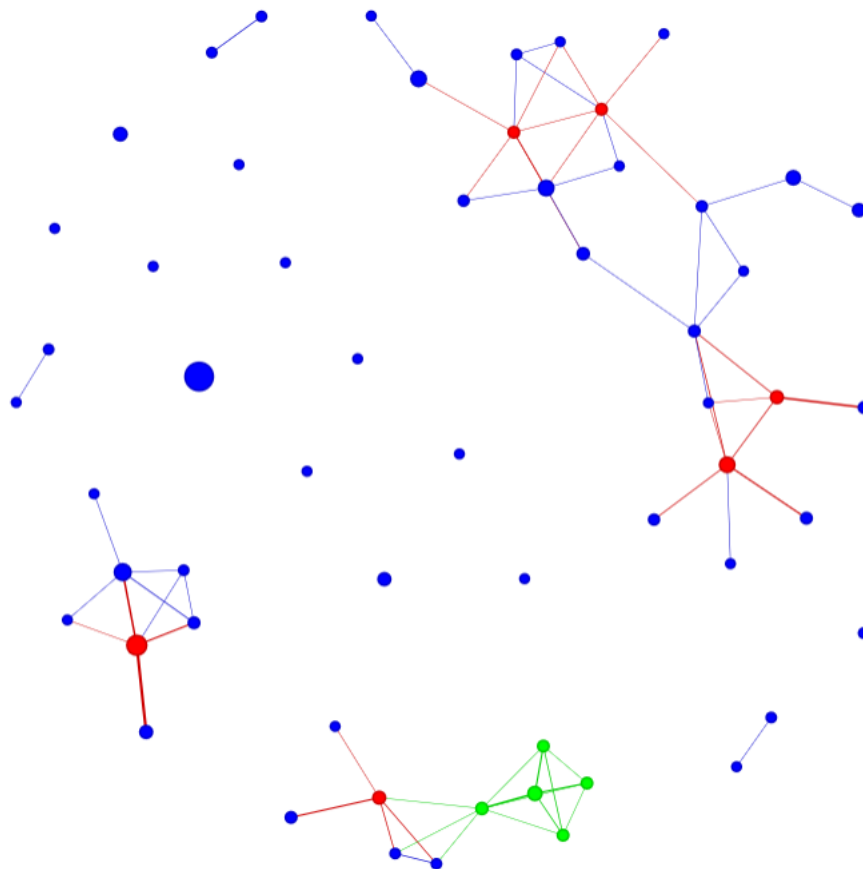


Figure 5: Réseau partiel de co-auteurs de l'unité MIA - PS coloré par appartenance aux groupes

## Perspectives et avenir du package

- Création et optimisation des fonctions permettant la création d'un réseau à partir des données
- Création et optimisation des fonctions permettant la création d'un modèle à partir du réseau obtenu

**Merci de m'avoir écouté**

**sources**

<https://www.ouvrirelascience.fr/plan-national-pour-la-science-ouverte/>