

Analyse de réseau, vocabulaire et aspects disciplinaires

Laurent Beauguitte
beaugittelaurent@hotmail.com

CNRS - UMR IDEES

19 janvier 2015

Objectifs de la séance

- connaître le contenu du cours, des TD et les modalités d'évaluation
- différencier analyse de réseau et analyse d'un réseau
- connaître le vocabulaire de base
- connaître les principales utilisations de l'analyse de réseaux selon les époques et les disciplines

Calendrier

- 19/01 : Introduction
- 20/01 : Analyse des réseaux sociaux / Réseaux spatiaux et analyse de flux
- 30/01 : Réseaux complexes / TD1 : R (igraph, statnet, tnet)
- 09/02 : Modélisations statistiques / TD2 : R (analyse de flux et cartographie)
- 13/02 : Visualisation / TD3 : Cytoscape et Gephi / exposés

Diplomatic dataset : <http://www.correlatesofwar.org/> : relations diplomatiques au niveau mondial de 1848 à 2006

- Diplomatic_Exchange_2006v1.csv (12 Mo)
- Diplomatic_Exchange_2006v1_codebook.pdf

6 variables

- ccode1 - Correlates of War state number of country 1 (origine)
- ccode2 - Correlates of War state number of country 2 (destination)
- year - Year of observation
- DR_at_1 - Diplomatic representation level of side 2 at side 1 (5 modalités)
- DR_at_2 - Diplomatic representation level of side 1 at side 2 (5 modalités)
- DE - Any diplomatic exchange between side 1 and side 2 (présence/absence)
- version - Current version of the data set (inutile)

DR_at_1 et DR_at_2 : 5 modalités (à simplifier)

0=no evidence of diplomatic exchange

1=chargé d'affaires

2=minister

3=ambassador

9=other

```
#exemple de script minimal

data <- read.table("Diplomatic_Exchange_2006v1.csv",
                  header = TRUE, sep=",")

head(data,3)
str(data)

#selection origine - destination - echange en 1990

data90 <- subset(data[,c(1,2,6)], data$year== "1990"
& data$DE == "1")
head(data90)

#transformation (fausse...) en graphe

library(igraph)
g90 <- graph.data.frame(data90)
```

Pistes possibles

- évolution après fin URSS ; (ex-)Yougoslavie avant et après 1995
- centrer l'analyse sur une région (Moyen-Orient, Amérique latine), une catégorie (ex. PMA) ou une comparaison d'espaces
- différencier liens forts, liens faibles
- expliquer présence des liens forts par d'autres variables (commerce, poids, absence de conflits etc.)

Ressources : COW datasets, packages cshapes et countrycode

Autre jeu de données : recommandations au Conseil des droits de l'homme

<http://www.upr-info.org/en/review>

7 variables : SuR (State under review); Recommendation (titre); RS (Recommending State); Response (Accepted or Noted); A (Action 5 modalités); Issue; Cycle (1 ou 2).

Pistes : comparer les 2 cycles; comparer graphe des recommandations Accepted et Noted

Extrait de la base

Displaying results 1 to 277 of 277 total results

SuR	Recommendation	RS	Response	A	Issue	C
Albania  EEG OIC, OIF	Sign and ratify the Optional Protocol to the ICESCR	Spain  WEOG EU, OEI	Accepted	5	<ul style="list-style-type: none"> ESC rights - general International instruments 	2
Albania  EEG OIC, OIF	Ratify the Optional Protocol to ICESCR	Portugal  WEOG EU, OEI	Accepted	5	<ul style="list-style-type: none"> ESC rights - general International instruments 	2
Albania  EEG OIC, OIF	Ratify the Optional Protocol to ICESCR	Israel  WEOG	Accepted	5	<ul style="list-style-type: none"> ESC rights - general International instruments 	2
Albania  EEG OIC, OIF	Ratify the Optional Protocol of the CRPD	Austria  WEOG EU	Accepted	5	<ul style="list-style-type: none"> Disabilities International instruments 	2
Albania  EEG OIC, OIF	Ratify the Optional Protocol of the CRPD	Israel  WEOG	Accepted	5	<ul style="list-style-type: none"> Disabilities International instruments 	2

Grappe valué Etat - Etat

Trois définitions de l'analyse de réseau

Analyse de réseau \neq analyse d'un réseau

Différentes disciplines, différentes méthodes et différents types de réseaux

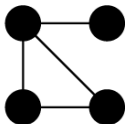
Un graphe = des points (ensemble fini non vide) et des lignes (ensemble fini éventuellement vide) : $G(V, E)$.

Série d'oppositions :

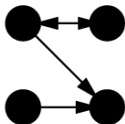
- orienté vs non orienté
- vide vs complet
- planaire vs non planaire
- simple vs pseudo (boucle - *loop* ou *self-tie*) ou multigraphe
- booléen vs valué ($G(V, E, W)$)
- connexe vs non connexe
- *one-mode* vs *two-mode* (biparti)

Types de réseaux

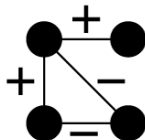
Simple, non orienté
(*graph*)



Simple, orienté
(*directed graph - digraph*)

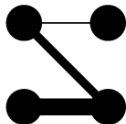


Simple, signé
(*signed graph*)

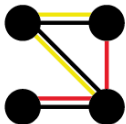


Types de réseaux

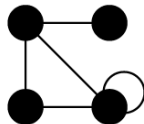
Simple, valué
(valued/weighted graph)



Multigraphe
(multigraph)

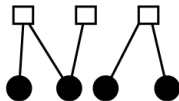


Avec boucle
(loop, self-tie)

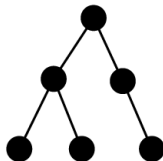


Types de réseaux

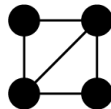
2-mode, biparti



Arbre

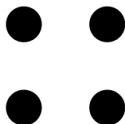


Planaire

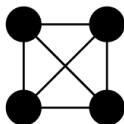


Types de réseaux

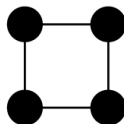
Vide



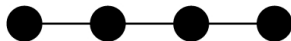
Complet



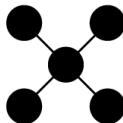
Cercle



Ligne



Étoile (*star*)



Vocabulaire de base

- chemin ou chaîne
- circuit ou cycle
- diamètre
- composantes connexes
- points d'articulation (*cut points*)
- isthmes (*bridges*)

3 représentations : graphe, matrice d'adjacence de taille $V \times V$, liste de liens (*edgelist*)

Un type particulier : le réseau biparti.

Exemple des réseaux discursifs à l'ONU

M. Wolfe (Jamaïque) (*parle en anglais*) : Ma délégation s'associe aux déclarations faites respectivement par le représentant du Guyana au nom des 14 membres de la Communauté des Caraïbes (CARICOM) et, à l'instant, par l'Ambassadeur de Tanzanie au nom des États membres du Groupe des États d'Afrique.

Recueil et mise en forme des données (matrice $N \times M$)

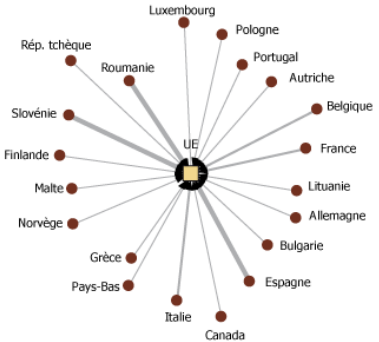
SESSION 45 (PV 1 à 82)

	G77	PMA	CE
Brésil	1	0	0
Ethiopie	1	1	0
France	0	0	1
Grèce	0	0	1
Irlande	0	0	3
Luxembourg	0	0	1
Pays-Bas	0	0	1
Portugal	0	0	1
RFA	0	0	1
Royaume-Uni	0	0	1

Visualisation (puis analyse)

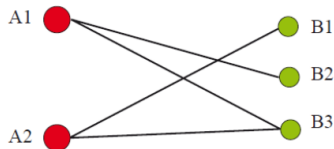


Session 51 (1996-1997)



L. Beauguitte, 2011.
Réalisé avec le logiciel Netdraw.

Très souvent, passage au réseau one-mode (valué - multiplication de la matrice par sa transposée ou inversement) puis choix d'un seuil



	A1	A2
A1	3	1
A2	1	3

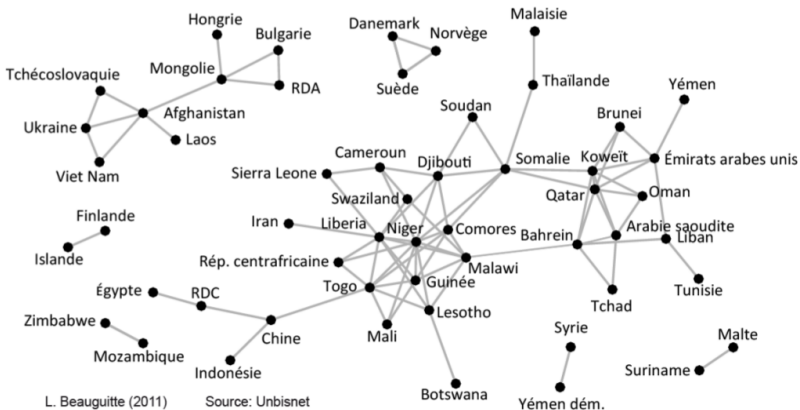
	B1	B2	B3
A1	0	1	1
A2	1	0	1

	B1	B2	B3
B1	2	0	1
B2	0	2	1
B3	1	1	2

Matrice de départ : État - résolution

Matrice valuée État - État puis seuillage à 60 %.

Figure 6 : Les parrainages communs dans 60% minimum des cas à la troisième Commission en 1986-1987.



4 grandes questions concernant les réseaux (plus ou moins selon les disciplines)

- comment caractériser le réseau ?
- comment caractériser les sommets (plus rarement les liens) ?
- quels sont les sous graphes pertinents ?
- comment expliquer la structure et l'évolution du graphe ?

Les ancêtres de la Social network analysis (SNA)

- Georg Simmel, l'ancêtre respectable (au rôle très limité)
- Les pionniers oubliés
- Jacob Moreno, l'ancêtre embarrassant
- Werner, Mayo, Gardner et Davis, Bavelas etc.

L'inventeur des matrices 2-mode (Hobson, 1858-1954) selon Freeman

	De Beers.	Premier.	Rand Mines.	Goldfields.	Chartered Company.
Beit, A.	I	—	I	—	I
Jameson, L. S. ...	I	—	—	—	I
Maguire, R. ...	—	—	—	I	I
Michell, Sir L. L.	I	—	—	—	I
Neumann, S. ...	—	I	I	—	—
Wernher	I	—	I	—	—

Figure 2.4. Hobson's two mode data



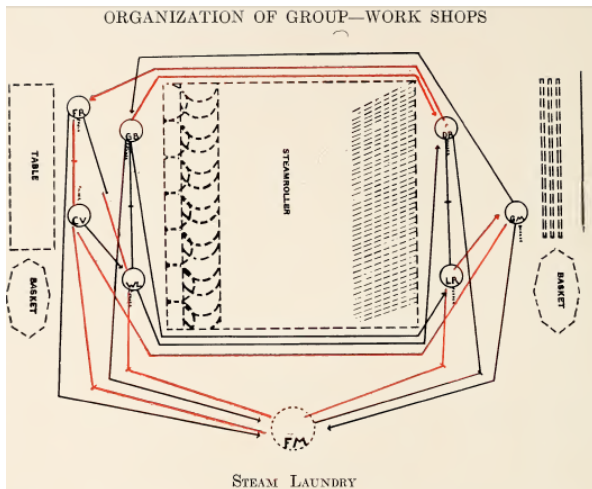


Fig. 1. 7 workers and 1 forewoman. Stella DR and Philamina LR, the feeders, reject each other. Hilda GR and Myrtle WL, the catchers, reject each other. Myrtle rejects the feeder opposite her, Philamina. Lillian FR and Rosalie CV, the two folders, attract each other. Lillian and Rosalie reject Myrtle. Esther GM, the shaker, is attracted to Lillian and rejects Hilda. Esther, Stella, Hilda, and Rosalie reject the fore-

Moreno, 1934, p. 133 - mais la *sociometry* s'écroule...

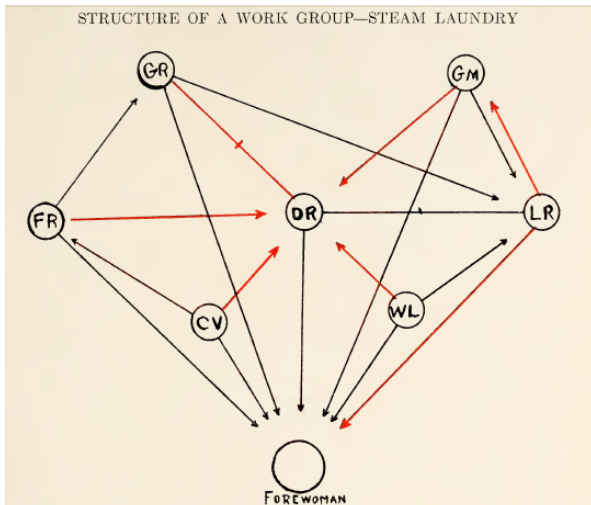


Fig. 2. Stella DR has gained a position of greater influence. In Fig. 1 she is the object of one attraction, Hilda GR; now she is the center of five attractions: Hilda GR, Myrtle WL, Lillian FR, Rosalie CV, Esther GM, and forms with Hilda a pair. She is rejected by Philamina LR whom she rejects in return. In Fig. 1 the forewoman is rejected by

Davis, Gardner et Gardner, 1941, *Deep South*.

Figure 1, showing which women attended each event, is reproduced from DGG (p. 148).

NAMES OF PARTICIPANTS OF GROUP I	CODE NUMBERS AND DATES OF SOCIAL EVENTS REPORTED IN <i>Old City Herald</i>													
	(1) 6/22	(2) 3/2	(3) 4/12	(4) 9/16	(5) 2/25	(6) 5/19	(7) 3/15	(8) 9/16	(9) 4/6	(10) 6/10	(11) 2/23	(12) 4/7	(13) 11/21	(14) 8/3
1. Mrs. Evelyn Jefferson.....	X	X	X	X	X	X		X	X					
2. Miss Laura Mandeville.....	X	X	X		X	X	X	X						
3. Miss Theresa Anderson.....		X	X		X	X	X	X	X					
4. Miss Brenda Rogers.....	X		X	X	X	X	X	X						
5. Miss Charlotte McDowd.....			X	X	X	X	X							
6. Miss Frances Anderson.....			X		X	X	X	X						
7. Miss Eleanor Nye.....				X	X	X	X	X						
8. Miss Pearl Ogleshorpe.....					X	X	X	X	X					
9. Miss Ruth DeSand.....					X	X	X	X						
10. Miss Verne Sanderson.....						X	X	X	X			X		
11. Miss Myra Liddell.....							X	X	X			X		
12. Miss Katherine Rogers.....							X	X	X	X		X	X	X
13. Mrs. Sylvia Avondale.....							X	X	X	X		X	X	X
14. Mrs. Nora Fayette.....						X	X	X	X	X	X	X	X	X
15. Mrs. Helen Lloyd.....							X	X		X	X	X		
16. Mrs. Dorothy Murchison.....							X	X	X		X			
17. Mrs. Olivia Carleton.....							X	X		X				
18. Mrs. Flora Price.....							X	X		X				

Figure 1. Participation of the Southern Women in Events

Mayo, Lloyd et Warner, années 30

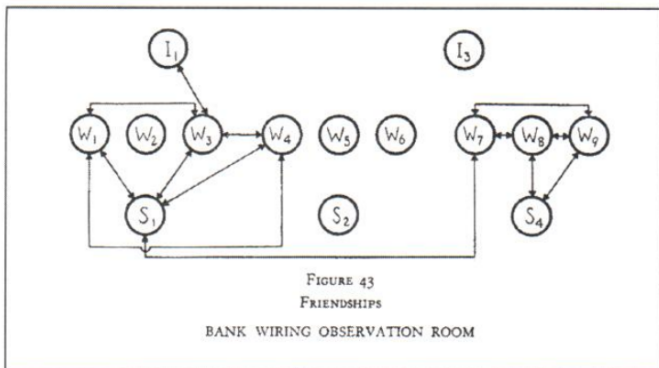


Figure 4.3. Friendships in the bank wiring room

Des précurseurs au paradigme : Harrison White et l'école de Harvard (cf cours 2)

- 1955 Thèse de physique au MIT
- 1960 Thèse de sociologie à Princeton
- 1965 Cours de licence : *An introduction to social relations* - direction de thèses dont Bonacich, Breiger, Granovetter, Levine, Wellman etc.*
- 1977 INSNA et *Connections*
- 1978 *Social Networks*
- 1981 *SunBelt*

* toujours vivant et influent (cf l'ouvrage de Duncan Watts, *Six degrees* et dédié à Strogatz et à Harrison White)

Analyse de réseaux en écologie : dès années 50, modélisation sous forme de graphe biparti

de Angelis, 1975

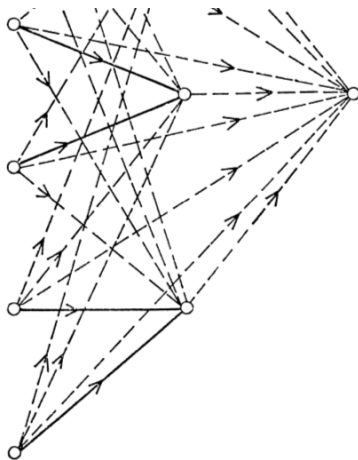


FIG. 1. Food web consisting of 10 species. Solid lines represent connections always assumed to exist; dashed lines represent potential connections, and the arrows denote the direction of biomass flow.

Armstrong, 1982

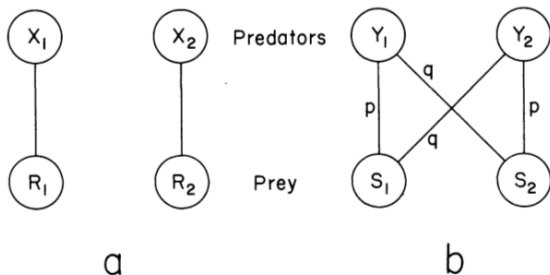


FIG. 1.—The model communities in question. *a*, The “unconnected” system (sec. 1, eqq. [1]). Each predator species feeds on only one prey species. *b*, The “connected” system (sec. 1.1, eqq. [6]). Each predator species feeds on both prey species. The letters p and q refer to the fractions of each prey’s productivity which flow up the “main links” and “cross links” of the system at equilibrium. Note that $p + q = 1$.

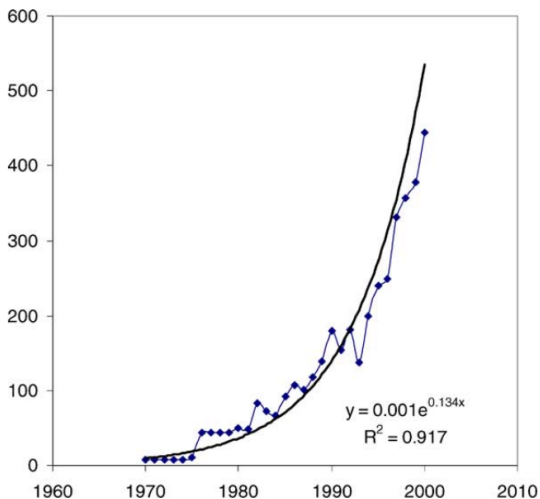
En géographie (cf cours 3) :

- années 40 : classification des réseaux hydrographiques (Horton-Strahler)
- années 60 : analyse des réseaux techniques et notamment de transport (Kansky, 1963)
- années 60 : modélisation des flux (Nystuen et Dacey, 1961)

Pas ou peu d'interactions avec les autres disciplines avant la fin des années 90 (idem pour la SNA ou l'écologie) : vocabulaire, logiciels, concepts différents.

Autres disciplines (histoire, économie, gestion...) : le plus souvent, importation des méthodes de la SNA.

Puis les physiciens arrivèrent... (voir cours 4)



Exponential growth of publications indexed by Sociological Abstracts containing 'social network' in the abstract or title (Borgatti, S. P. & Foster, P. C. (2003). The network paradigm in organizational research : A review and typology. *Journal of management*, 29(6), 991-1013)

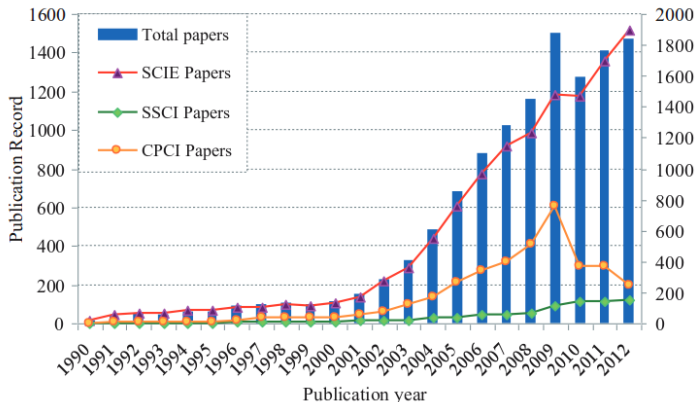


Fig. 1 Papers record indexed in WoS from 1990 to 2012

Articles contenant mot clef *complex network* (Fei-Cheng Ma, Peng-Hui Lyu and Xiao-Guang Wang, 2014, Knowledge Discovery of Complex Networks Research Literatures, Springer)

Des mondes séparés : small-world networks en SNA et en physique (Freeman, 2001)



12 avril 2013 : 46 *packages* contenant le terme *network* dans leur titre ou leur description sur CRAN

13 janvier 2015 : plus de 65 *packages* contenant le terme *network*

NetComp	Network Generation and Comparison
NetData	Network Data for McFarland's SNA R labs
NetIndices	Estimating network indices, including trophic structure of foodwebs in R
NetPreProc	NetPreProc: Network Pre-Processing and normalization
nets	Network Estimation for Time Series
network	Classes for Relational Data
NetworkAnalysis	Statistical inference on populations of weighted or unweighted networks
networkDynamic	Dynamic Extensions for Network Objects
networksis	Simulate bipartite graphs with fixed marginals through sequential importance sampling
networkTomography	Tools for network tomography

Cinq grandes catégories

- les généralistes
- les spécialistes
- les adaptables
- les inutiles & inutilisables
- les morts

Packages généralistes

Deux *packages* généralistes et non compatibles

`igraph`

package développé par des informaticiens & des physiciens
très nombreuses mesures issues de la sociologie et de la physique
objet `igraph`

`statnet`

ensemble de *packages* développé par des sociologues
nombreuses mesures issues de la sociologie et de la physique et modèles
statistiques (*ERGM - Exponential Random Graph Model*)
objet `network`

Dans les deux cas, site internet, documentation, tutoriaux et *mailing list*

Lequel choisir ?

igraph

importation des données plus simple

résultats présentés de façon parfois peu lisibles

plus d'indicateurs disponibles

sortie graphique par défaut laide & fausse 3D

statnet

(un peu) moins d'indicateurs

importation moins aisée (attributs)

modèles statistiques

vraie 3D

Utilisez les deux. . .

Un exemple : propriétés du graphe

statnet

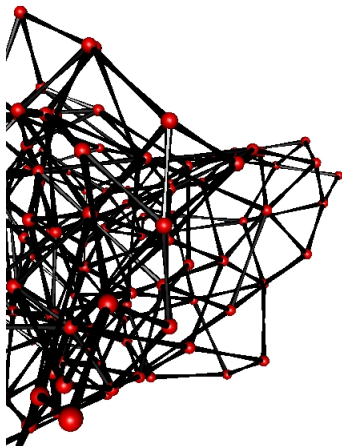
```
gMobR

## Network attributes:
##   vertices = 138
##   directed = TRUE
##   hyper = FALSE
##   loops = FALSE
##   multiple = FALSE
##   bipartite = FALSE
##   total edges= 136
##     missing edges= 0
##     non-missing edges= 136
##
## Vertex attribute names:
##   vertex.names
```

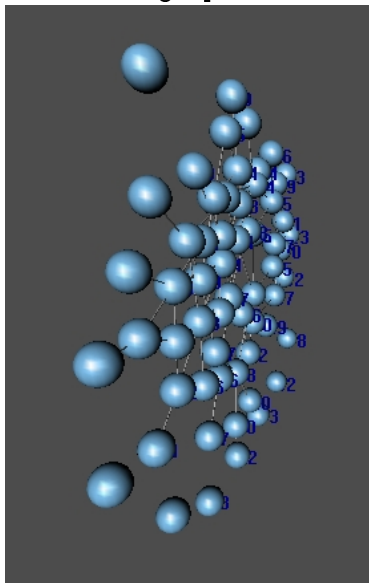
igraph

```
## IGRAPH DN-- 138 136 --
## attr: name (v/c)
```

statnet



igraph



Packages spécialisés

- *package* QuACN : analyse de réseaux complexes et biologiques (mesures très spécifiques)
- *package* bipartite : analyse de réseaux bipartis, développé par des écologues
- *package* enaR : analyse de réseaux écologiques (procédures MatLab)
- *package* tnet : analyse de réseaux valués (simples et bipartis)
- *package* egonet : analyse de réseaux personnels
- *package* intergraph : transforme les objets `network` (statnet) en objet `igraph` et vice versa (mais pas les graphes bipartis...)

Packages adaptables

Packages non conçus pour l'analyse de réseaux mais qu'il est possible de détourner...

Analyse de réseau -> calcul matriciel

Packages inutiles & inutilisables

Inutilisables : documentation incomplète et/ou incompréhensible, pas de mise à jour (`blockmodeling` - 2010-01-13)

Inutiles : non multiplateforme, maintenance non assurée (`InteractiveIgraph`)

Note

This project is not fully completed, but unfortunately it is not in active development any more. Up till now it fulfills all my needs. In the case of new ideas i might join further development.

Packages morts

Un des vrais problèmes concernant la portabilité de R

Packages récemment disparus : biGraph, Rgraphviz, graphComp

Capacité à maintenir le *package* en fonction des changements de version de R et des *packages* parents

Package 'biGraph' was removed from the CRAN repository.

Formerly available versions can be obtained from the [archive](#).

Package is broken since a (preliminary announced) API change in package igraph,

Inconvénients

Packages différents utilisant des objets différents. Difficile de jongler entre eux.

`igraph` : objet `igraph`; `statnet` : objet `network`; `tnet` : objet `matrix`
etc.

Pas de *package* dédié à l'analyse des graphes planaires!

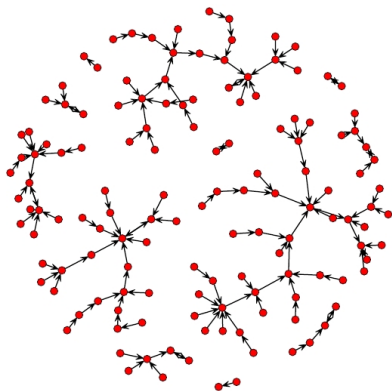
Mortalité élevée des *packages*

Visualisation correcte pour du travail exploratoire

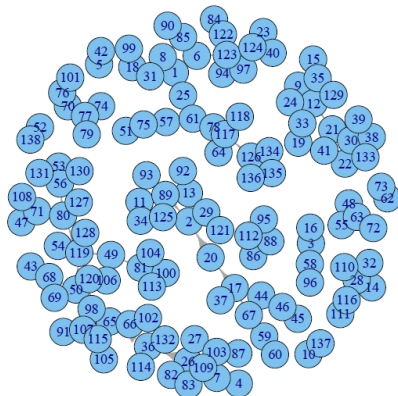
Visualisation insuffisante comme outil de communication

Visualisation par défaut

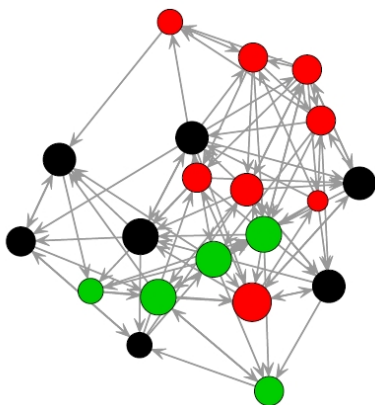
statnet



igraph



Visualisation avec (petit) script



```
gplot(g, displayisolates=FALSE,  
      vertex.cex=sqrt(degree(g, cmode="indegree")),  
      edge.col="grey60",  
      vertex.col=groupes,  
      mode="fruchtermanreingold")
```

<http://statnet.csde.washington.edu/>

<http://igraph.sourceforge.net/>

<http://groupefmr.hypotheses.org/>, voir notamment le billet « Analyse de réseaux avec R : des ressources gratuites pour un logiciel libre ! »