

# L'analyse des réseaux historiques

Concepts et principes fondamentaux

Sébastien de Valeriola, ULB  
[Sebastien.De.Valeriola@ulb.be](mailto:Sebastien.De.Valeriola@ulb.be)

Nicolas Ruffini-Ronzani, UNamur / Archives de l'État  
[nicolas.ruffini@unamur.be](mailto:nicolas.ruffini@unamur.be)

UCLouvain, 30 mars 2024

# Table des matières

## **A. INTRODUCTION**

1. L'analyse des réseaux sociaux : présentation générale
2. Les graphes
3. Manipulation et visualisation
4. Analyse
5. Quelques exemples d'utilisation



HUMANITÉS NUMÉRIQUES :

DE NOUVEAUX OUTILS POUR LE MÉDIÉVISTE



# Atelier n° 1

## Initiation à l'analyse de réseaux

- R.M.B.L.F. -

*Réseau des Médiévistes Belges de Langue Française*



Louvain-la-Neuve, 8 mai 2018

Nicoas Ruffini-Ronzani (UNamur / UVSQ) et Sébastien de Valeriola (UCLouvain / UGent)

nicolas.ruffini@unamur.be

sebastien.devaleriola@uclouvain.be

Un *vade mecum* en libre accès : <https://paths.unamur.be/prame/cartulaires-de-wallonie/gephi>

# **1. L'ANALYSE DES RÉSEAUX SOCIAUX :**

## **PRÉSENTATION GÉNÉRALE**

# L'analyse des réseaux sociaux

Le principe de l'analyse des réseaux sociaux est simple : il s'agit d'étudier un groupe d'entités sur base des **relations** qu'elles entretiennent entre elles.

Ces entités et ces relations peuvent représenter des réalités très différentes :

- des personnes qui échangent des lettres ;
- des marchands qui concluent un contrat ensemble ;
- des villes qui sont reliées entre elles par des routes ;
- des villages qui partagent un « champ de vision » similaire ;
- des manuscrits qui contiennent les mêmes textes ;
- des mots qui partagent une certaine similarité sémantique ;
- ...

L'étape initiale de définition même du réseau analysé est cruciale : **quelles sont les entités** que nous étudions, et **quelles relations** entre elles considérons-nous ?

Des choix différents (même légèrement) mènent à des réseaux différents.

# L'analyse des réseaux sociaux

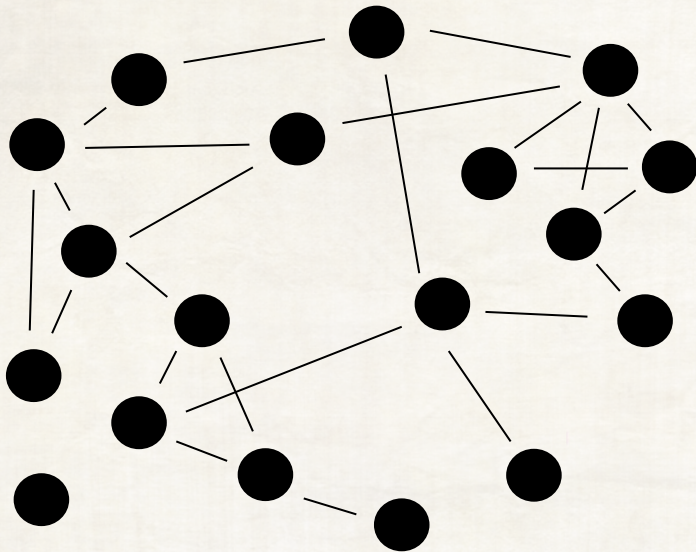
L'engouement prononcé de **l'historiographie** pour cet ensemble de techniques est plutôt récent :

- dans les années 1950 et 1960, les sociologues s'intéressent aux données relationnelles, dans un contexte fort différent de celui de l'analyse des réseaux sociaux :
  - présence conjointe à un même événement ;
  - partage de certaines caractéristiques comme l'origine, l'âge ;
  - etc.
- dès les **années 1970**, certains historiens commencent à s'intéresser de près aux relations entre les individus ; cet intérêt reste cependant marginal, peut-être en raison des lourdes manipulations de prétraitement à effectuer sur les données historiques (qui sont généralement moins lourdes pour des données sociologiques) ;
- au cours des **années 1990**, l'analyse des réseaux sociaux commence à se diffuser plus largement au sein des études historiques, avant de s'imposer comme une méthode à part entière dans les **années 2000 et 2010**.

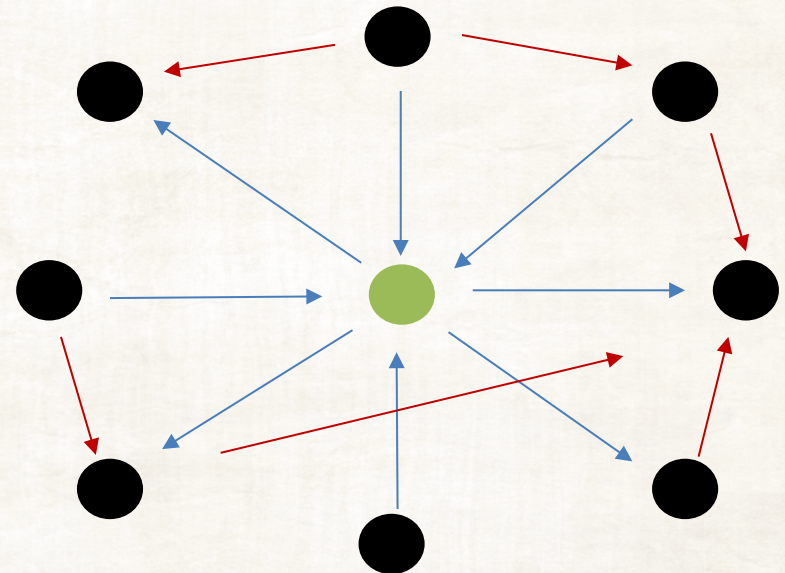
# Deux types d'analyses

On distingue deux types d'analyse, qui témoignent de deux approches différentes de la relation sociale :

La démarche *globale* (ou *complète*) tente de capturer les **relations essentielles** qui lient les individus appartenant à un groupe social, et considère donc **tous les membres** de celui-ci, et toutes les dépendances entre eux.



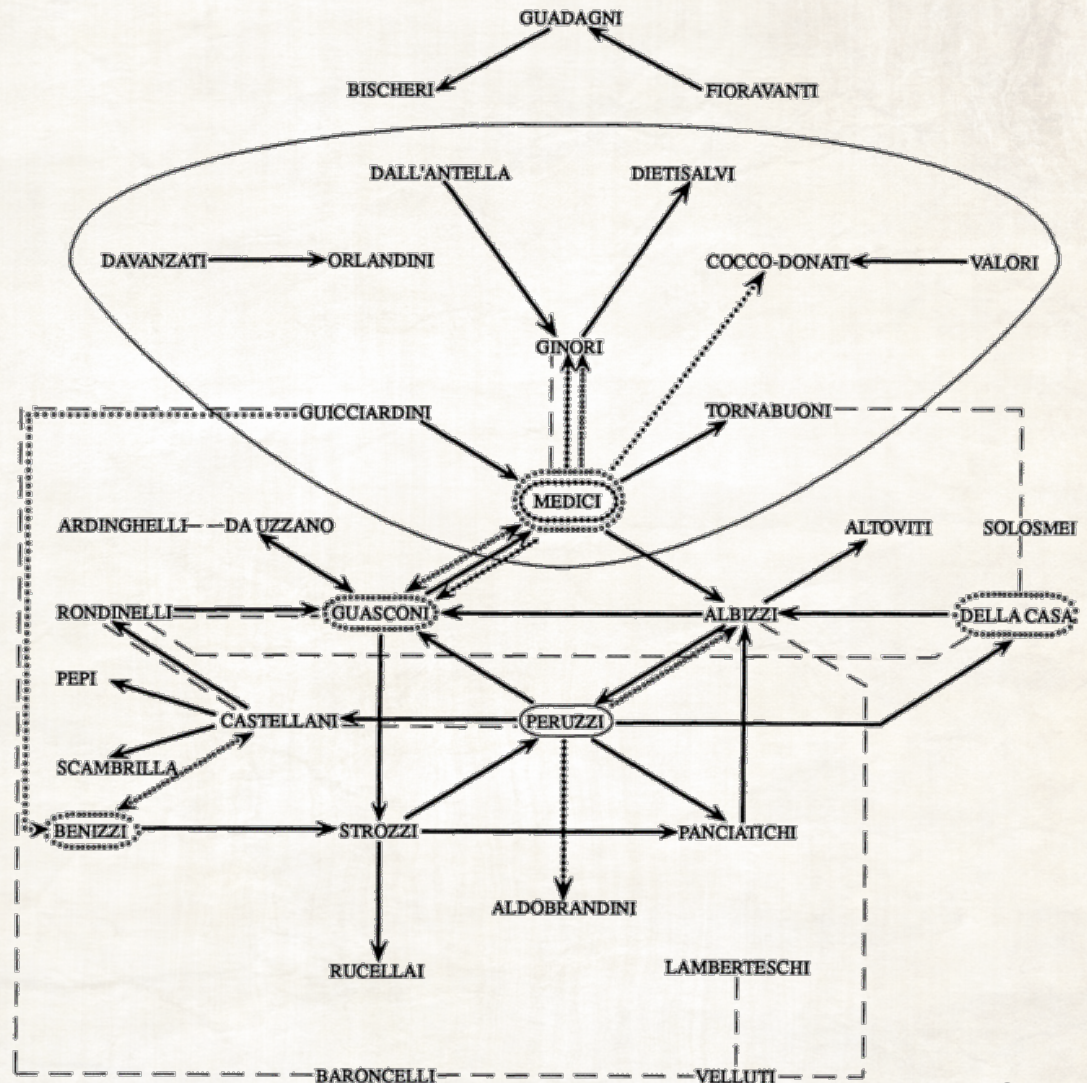
La démarche *égocentrique* (ou *personnelle*) prend le point de vue inverse, et étudie les différents réseaux **d'un individu en particulier**.



## Exemple de démarche globale

John Padgett et Christopher Ansell analysent les relations de mariage, d'affaires et de voisinage au sein de la communauté aristocrate florentine du xv<sup>e</sup> siècle. Ils parviennent à identifier des groupes sociaux, soudés de différentes manières, qui correspondent à des clans oligarchiques distincts.

ANSELL, C. ET PADGETT, J., « Robust action and the rise of the Medicis, 1400-1434 », in *American journal of sociology*, 98, n° 6, 1993, p. 1259-1319





## Exemple de démarche égocentrique

Margaret Mullett extrait des lettres de l'archevêque Théophylacte d'Ochrid toutes sortes des renseignements qui lui permettent d'étudier les relations de l'ecclésiastique avec les dignitaires religieux et laïques du début du XII<sup>e</sup> siècle.



MULLETT, M., *Theophylact of Ochrid : Reading the letters of a Byzantine archbishop*, Aldershot, Ashgate, 1997.

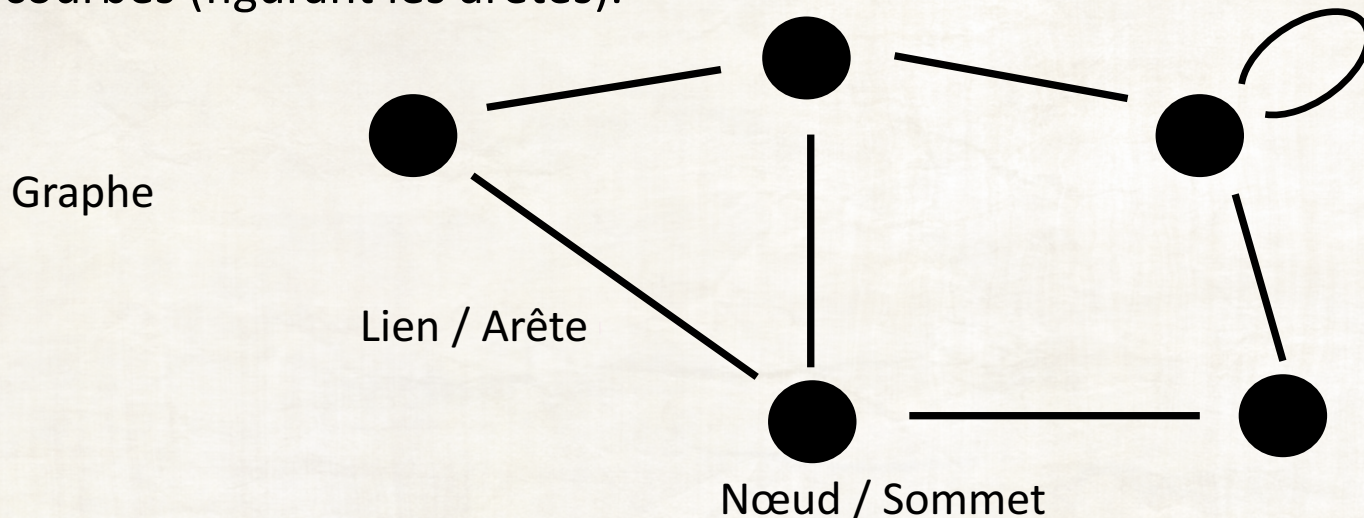
## **2. LES GRAPHES**

# Définition formelle

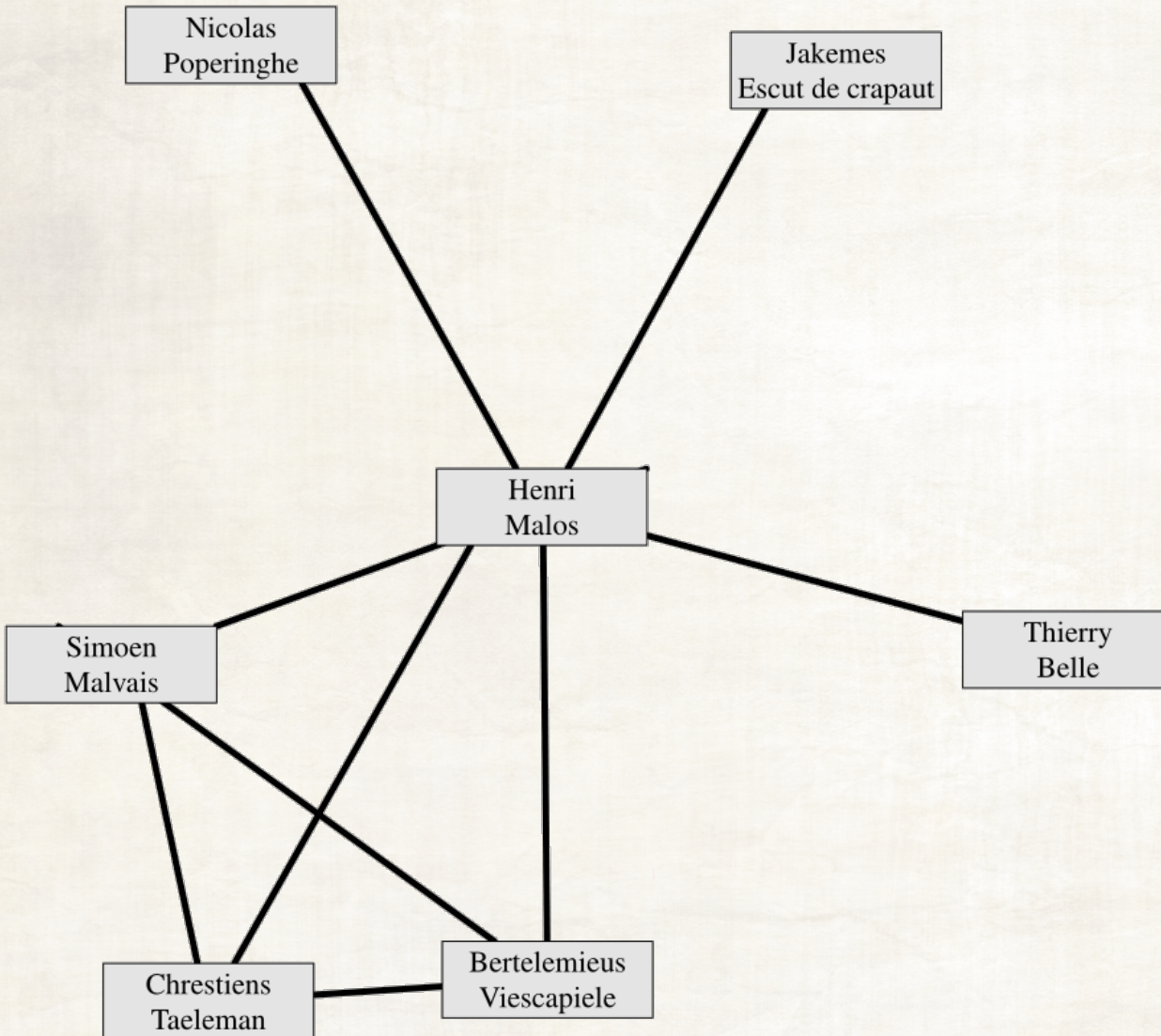
La base de l'analyse des réseaux est la représentation du réseau par un objet mathématique : le graphe.

Un *graphe* est un ensemble d'éléments appelés *sommets* accompagné d'un ensemble de relations entre ceux-ci, appelées *arêtes*.

Il est habituel de représenter graphiquement cette structure à l'aide de formes géométriques (figurant les sommets) reliées entre elles par des courbes (figurant les arêtes).



# Exemple de graphe



Nous pouvons ainsi par exemple considerer le réseau égocentré d'un acteur économique yprois du XIII<sup>e</sup> siècle (obtenu à partir d'un corpus de chirographes), et le graphe qui lui est associé.

Les sommets de ce graphe représentent des bourgeois yprois, qui sont joints deux à deux par une arête s'ils sont liés commercialement (dans un sens encore à définir).

Le graphe possède sept sommets et neuf arêtes.

# Graphes orientés et non-orientés

Les arêtes d'un réseau peuvent être « à double sens », auquel cas elles rendent compte de relations symétriques ; on parle alors de graphe *non orienté*.

C'est le cas du graphe représenté sur le slide précédent : les arêtes peuvent par exemple y figurer la relation « X apparaît avec Y dans un même chirographe ».

Celle-ci est bien symétrique : si le bourgeois Henri Malos apparaît dans un chirographe avec Jakemes Escut de crapaut, Jakemes Escut de crapaut apparaît évidemment dans un chirographe avec Henri Malos.

Au contraire, les arêtes d'un graphe *orienté* sont « à sens unique » et rendent compte de relations asymétriques.

La relation « être créancier de » est un exemple de telle relation : si Jakemes Escut de crapaut est le créancier de Henri Malos, il est incorrect de dire que Henri Malos est forcément le créancier de Jakemes Escut de crapaut.

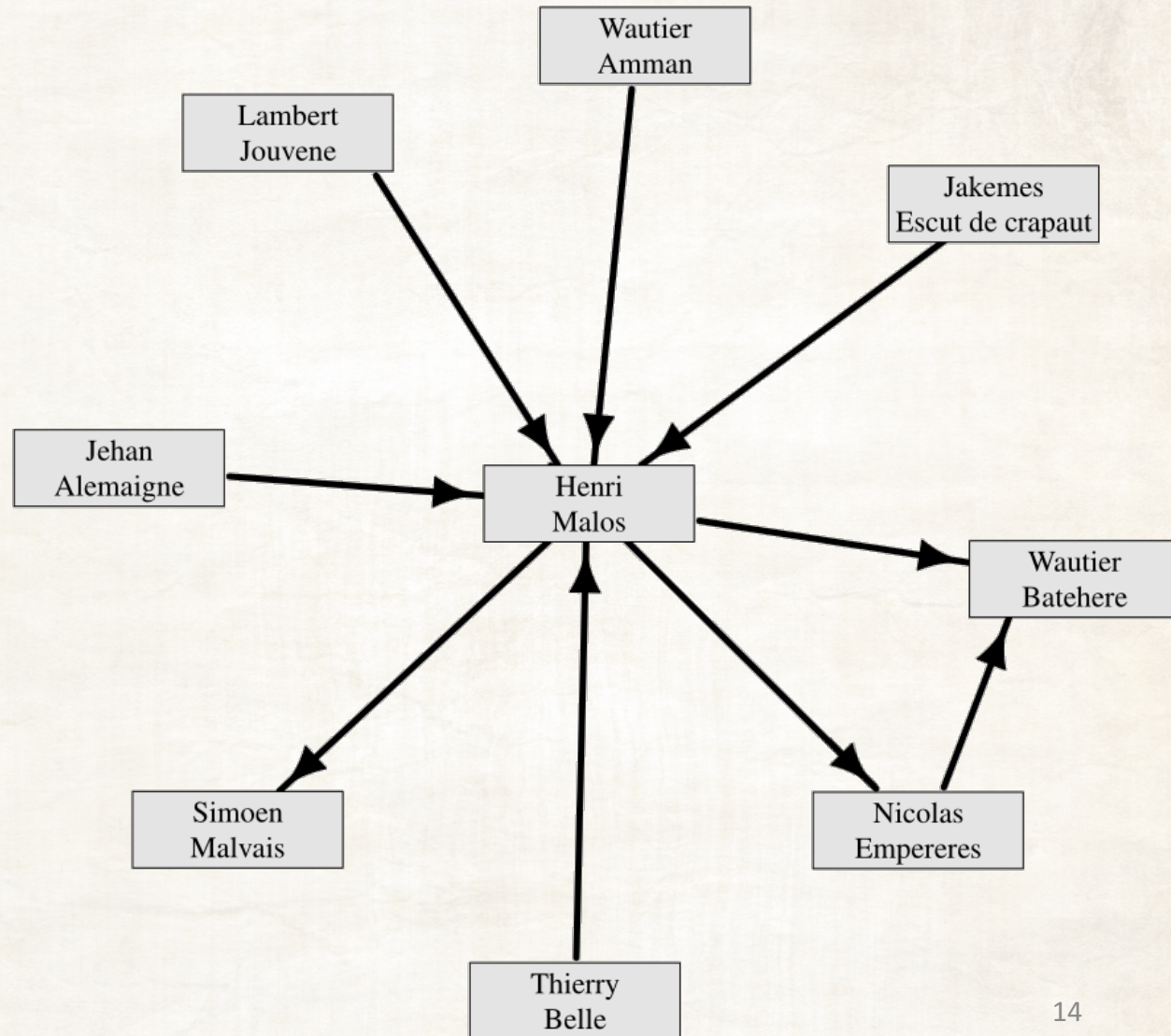
# Exemple de graphe orienté

Les arêtes sont alors représentées par des flèches.

La flèche reliant

Nicolas Empereres  
à  
Wautier Batehere

signifie donc qu'il existe un chirographe dans lequel le premier joue le rôle du créancier, tandis que le second joue le rôle du débiteur.



# Graphes pondérés

Il arrive que les relations entre deux personnes soient multiples.

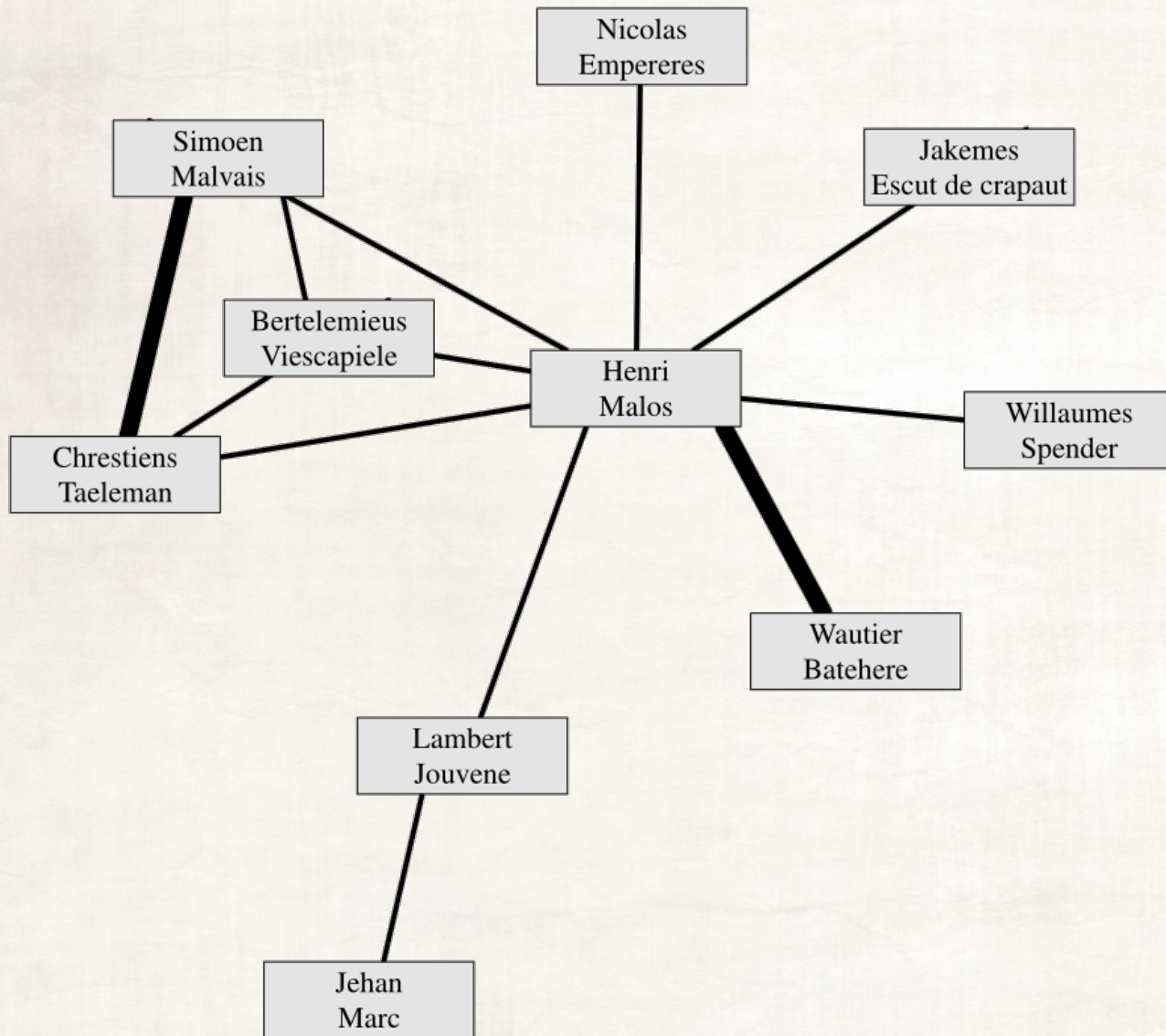
Dans les exemples précédents, il est envisageable que deux individus aient conclu plus d'un contrat ensemble.

Plutôt que de tracer plusieurs arêtes, on associe à chacune d'elles un nombre entier, appelé *poids*, qui représente le nombre de fois que la relation entre les deux sommets est attestée.

On parle alors de graphe *pondéré* ou *valué*.

Cette caractéristique est parfois figurée par l'affichage du poids à côté de l'arête, d'autres fois par la largeur de l'arête elle-même.

# Exemple de réseau pondéré



Toutes les arêtes ont un poids égal à 1, sauf celle qui relie

Simoen Malvais  
et  
Chrestiens Taelman

et celle qui relie

Henri Malos  
et  
Wautier Batehere

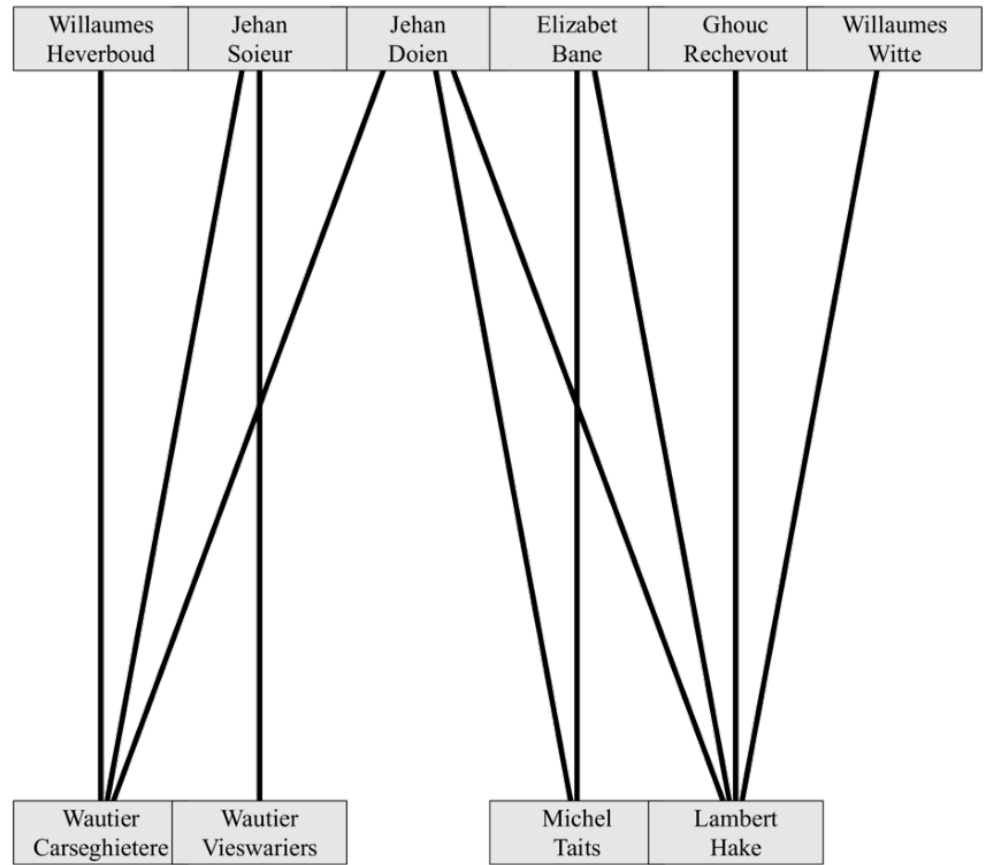
qui sont de poids 2.



# Graphes bipartis

Dans un graphe biparti, les sommets appartiennent à deux catégories, de telle sorte que les arêtes ne relient pas les sommets des mêmes catégories.

Par exemple, on pourrait définir les sommets comme soit des créanciers professionnels, soit des débiteurs, et des arêtes qui représentent la relation « prête de l'argent à ».



# **3. MANIPULATION ET VISUALISATION**

# Matrice d'adjacence

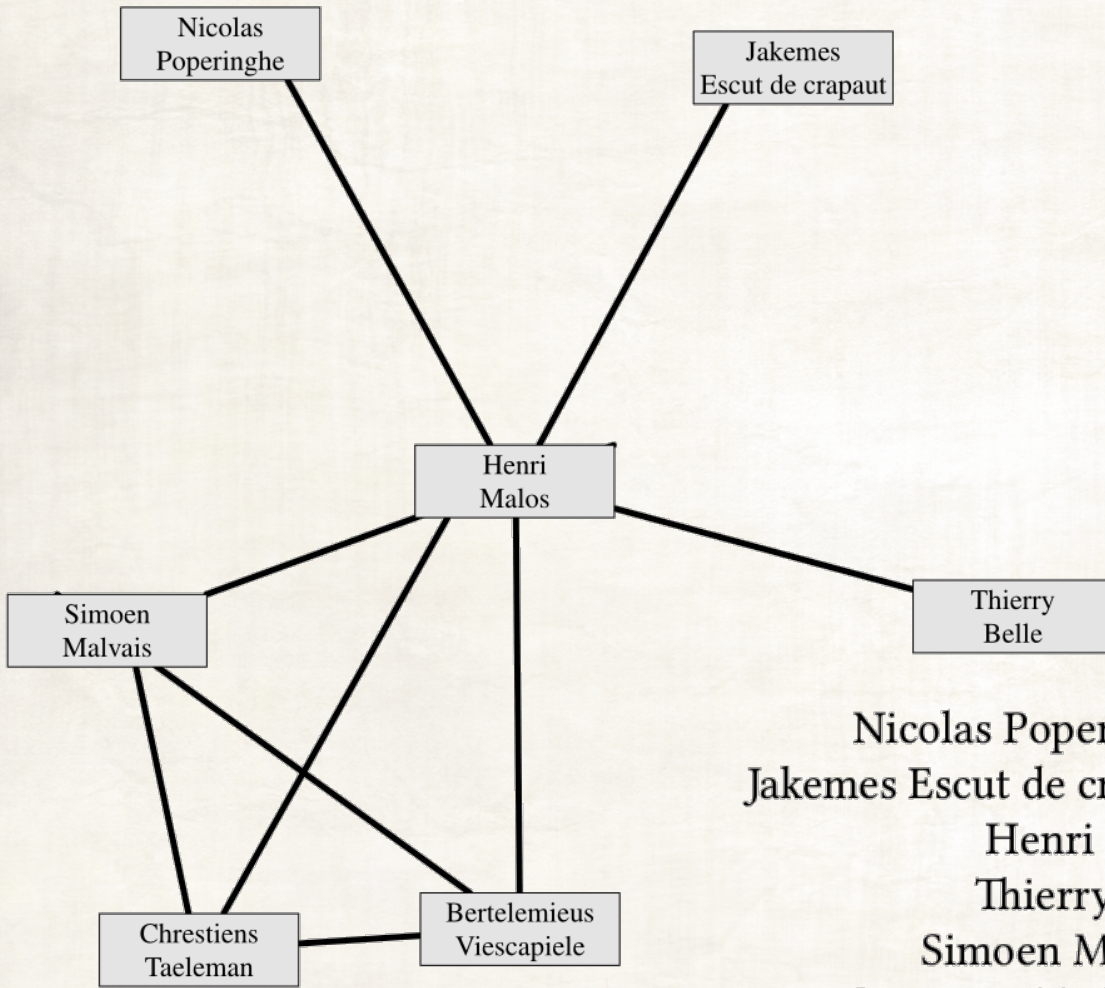
Les sommets et les arêtes d'un réseau constituent une structure mathématiquement pratique et efficace, mais trop complexe pour être manipulée telle quelle par un ordinateur.

Pour cette raison, on associe à chaque réseau une *matrice d'adjacence*.

Il s'agit d'un tableau carré ayant autant de lignes et de colonnes que le nombre total de sommets, et dont chacune des entrées vaut 1 si les sommets associés à la ligne et à la colonne de l'entrée sont connectés, ou 0 s'ils ne sont pas connectés :

$$\text{matrice d'adjacence}_{i,j} = \begin{cases} 1 & \text{s'il existe une arête reliant les sommets } i \text{ et } j, \\ 0 & \text{sinon.} \end{cases}$$

# Exemple de matrice d'adjacence



	Nicolas Poperinghe	Jakemes Escut de crapaut	Henri Malos	Thierry Belle	Simoen Malvais	Chrestiens Taelleman	Bertelemius Viescapiele
Nicolas Poperinghe	0	0	1	0	0	0	0
Jakemes Escut de crapaut	0	0	1	0	0	0	0
Henri Malos	1	1	0	1	1	1	1
Thierry Belle	0	0	1	0	0	0	0
Simoen Malvais	0	0	1	0	0	1	1
Chrestiens Taelleman	0	0	1	0	1	0	1
Bertelemius Viescapiele	0	0	1	0	1	1	0

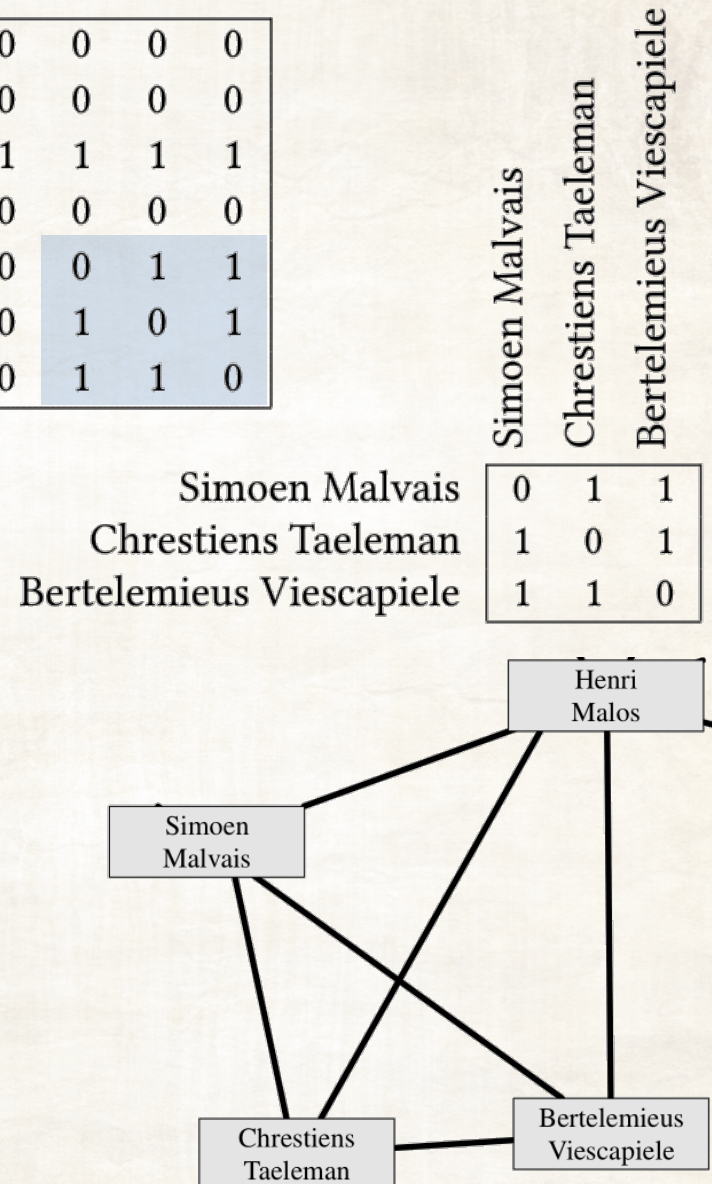
# Intérêt de la représentation matricielle

L'existence d'une représentation matricielle de chaque réseau n'est pas anodine : elle ouvre la porte à une multitude d'outils disponibles dans diverses branches des mathématiques.

0	0	1	0	0	0	0
0	0	1	0	0	0	0
1	1	0	1	1	1	1
0	0	1	0	0	0	0
0	0	1	0	0	1	1
0	0	1	0	1	0	1
0	0	1	0	1	1	0

Par exemple, certaines techniques permettent de repérer des sous-matrices particulières, comme celle qui est visible dans la matrice d'adjacence donnée en exemple.

Le sous-réseau correspondant possède en effet une caractéristique particulière, facilement observable dans ce modeste cas particulier, mais sans doute beaucoup plus difficile à reconnaître pour des réseaux plus importants : tous les sommets sont connectés à tous les autres (on parle d'une *clique*).



# Visualisation d'un graphe

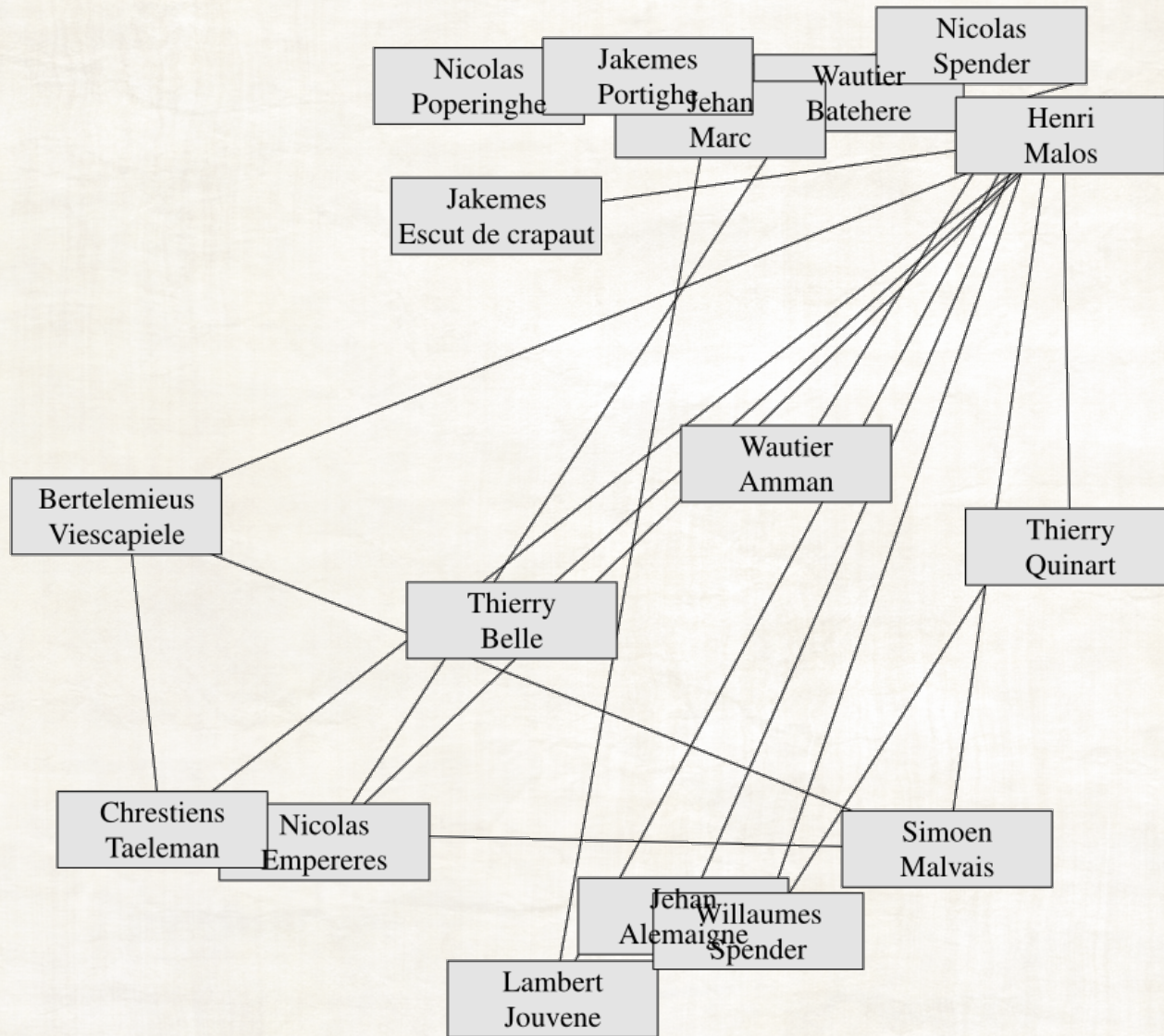
L'ordinateur est capable d'appliquer des méthodes quantitatives non triviales, mais ne possède évidemment aucune intuition par rapport au réseau qu'il traite.

Pour cette raison, les spécialistes de la théorie des réseaux ont développé des techniques permettant de visualiser les ensemble de sommets et leurs arêtes, afin de se faire une idée des directions dans lesquelles investiguer :

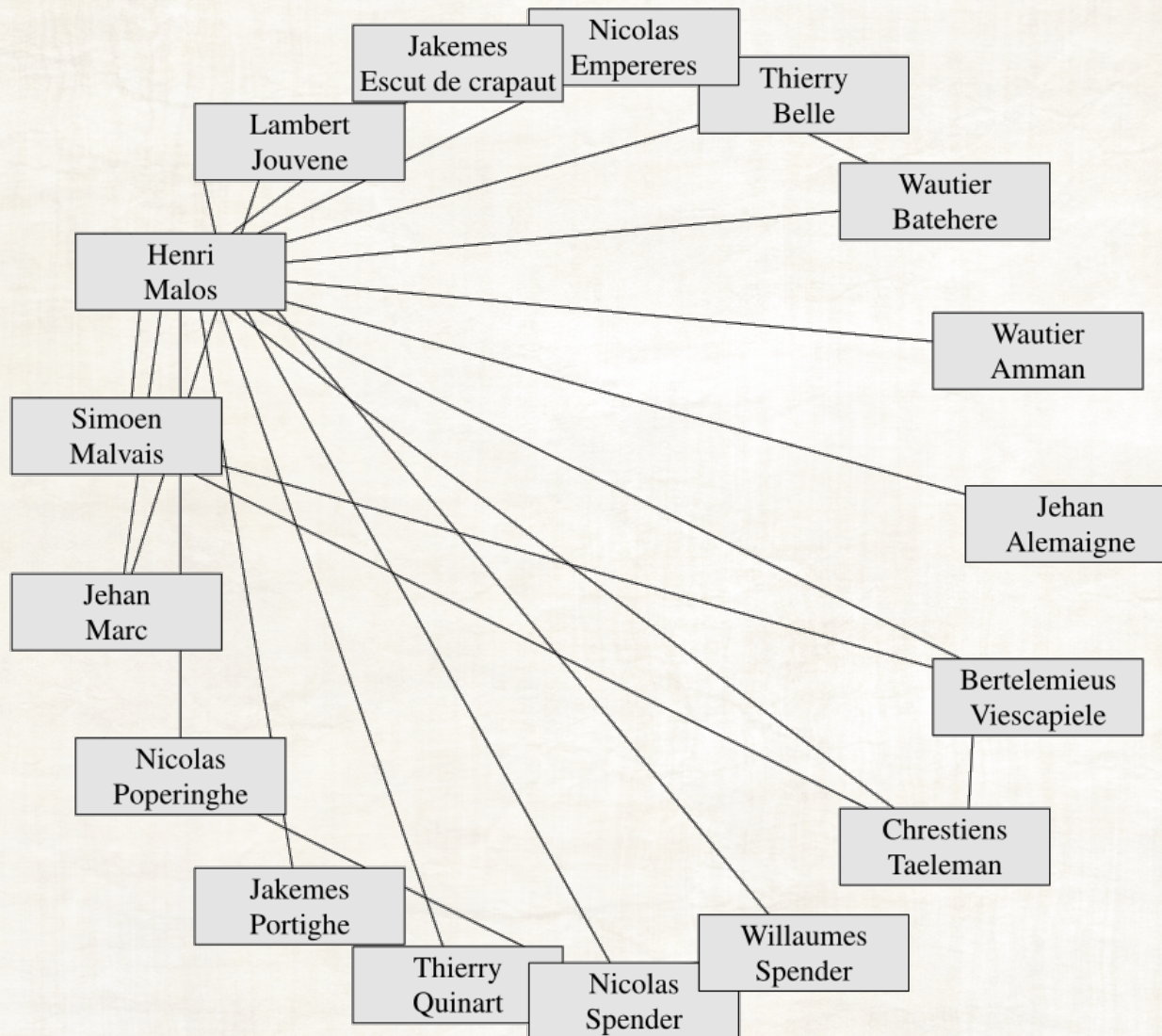
- le graphe possède-t-il une structure en étoile, suggérant que l'un des individus du réseau y occupe une place particulièrement importante ? ;
- est-il constitué de plusieurs sous- graphe faiblement liés entre eux ? ;
- etc.

C'est à ce genre de questions que l'analyste doit répondre, afin de décider l'orientation qu'il donnera à son analyse.

# Plusieurs visualisations d'un même graphe : disposition aléatoire

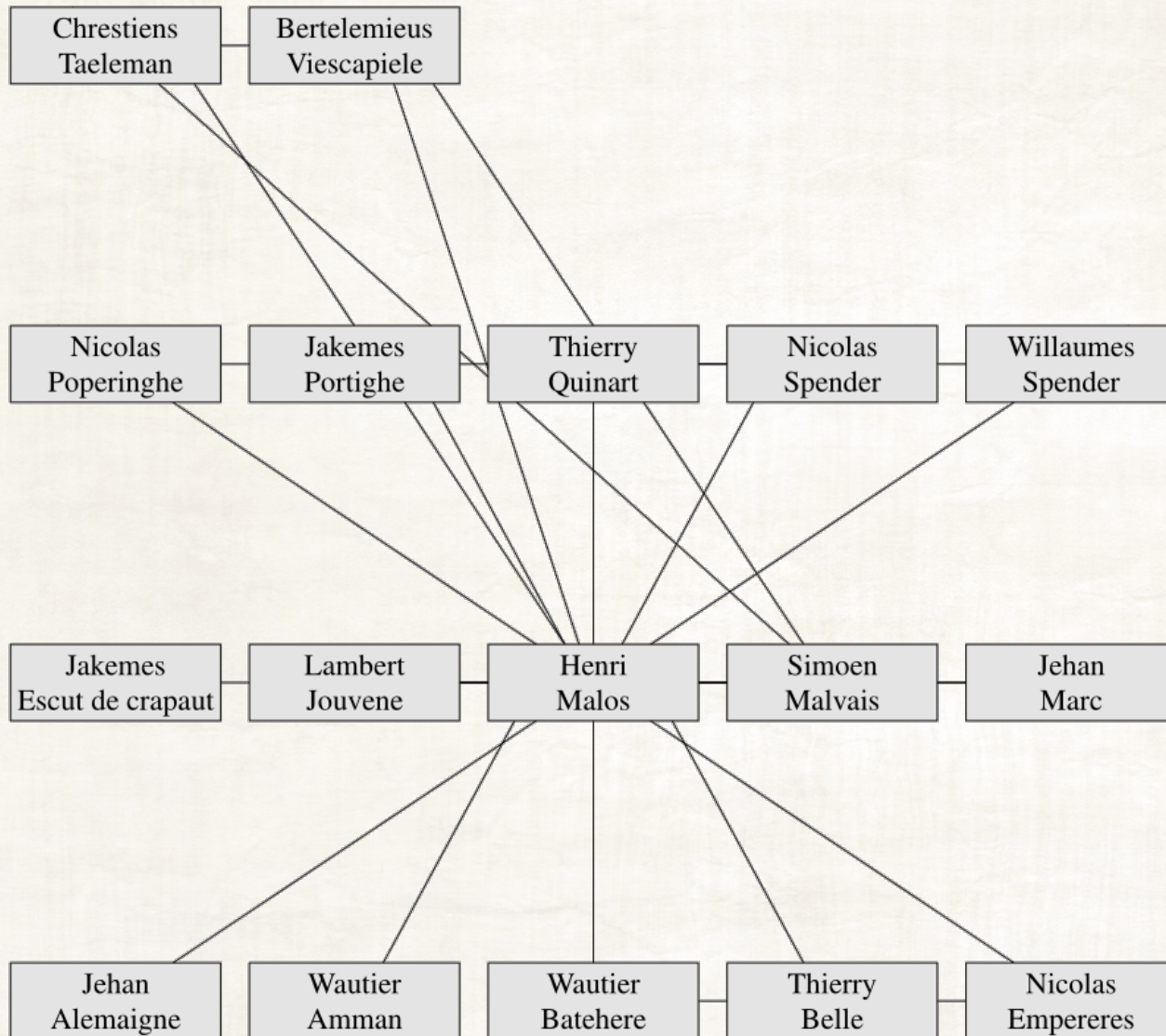


# Plusieurs visualisations d'un même graphe : disposition circulaire

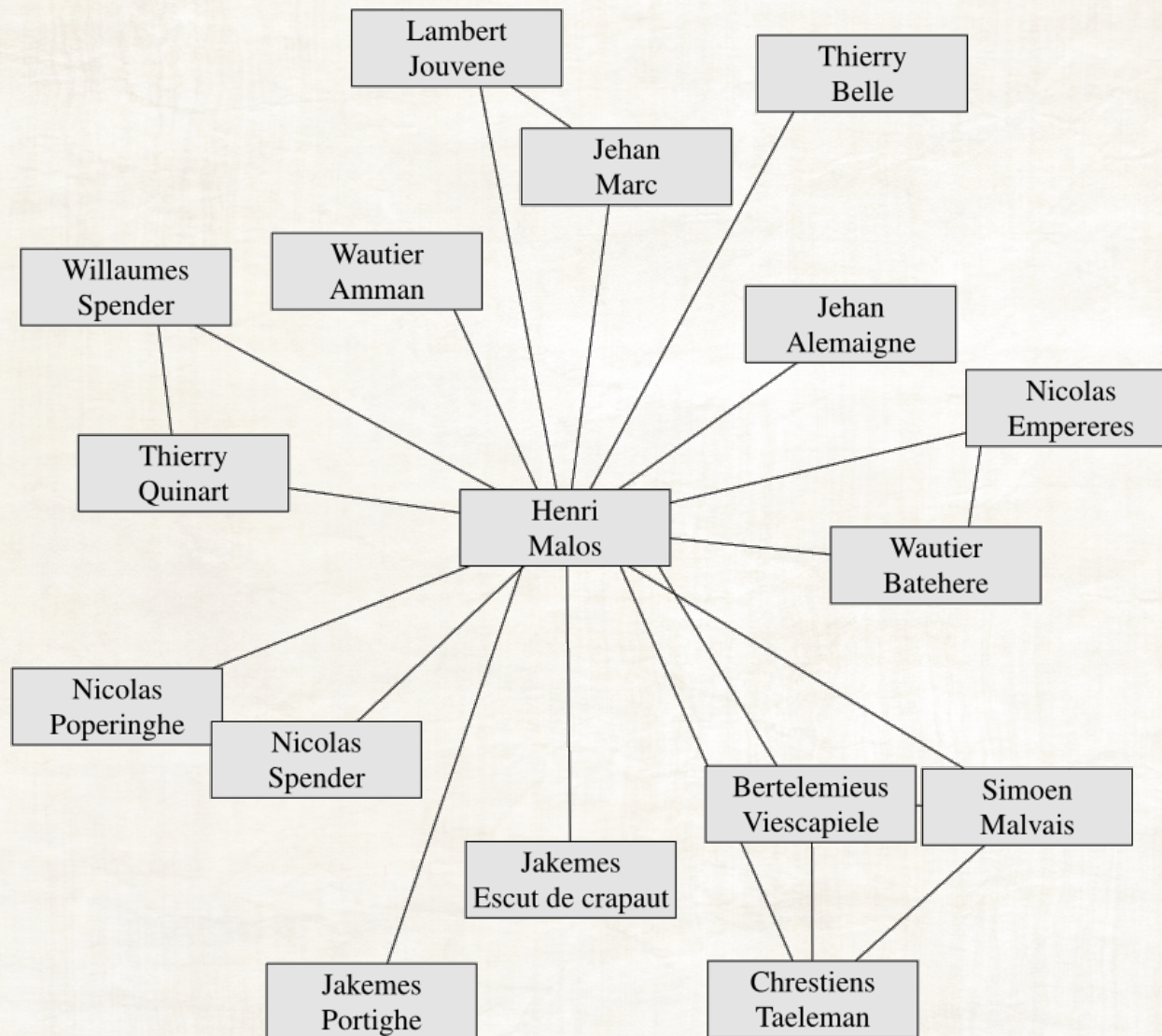




# Plusieurs visualisations d'un même graphe : disposition sur une grille



# Plusieurs visualisations d'un même graphe : disposition de Davidson-Harek

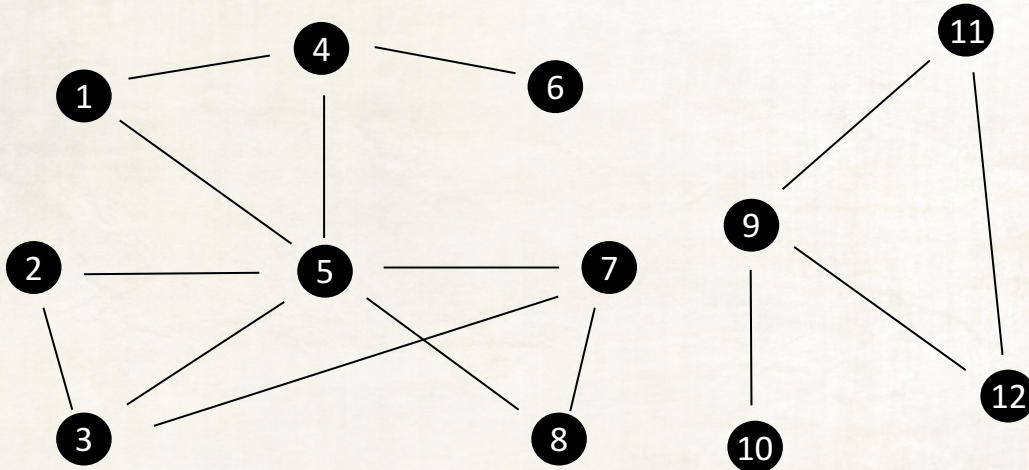


# **4. ANALYSE**

# Connexité et distance

Une première question à poser lors de l'analyse d'un graphe est sans doute celle de sa *connexité* : peut-on, en empruntant successivement des arêtes, voyager de n'importe quel sommet à n'importe quel autre ?

Si ce n'est pas le cas, on peut identifier les *composantes connexes* du graphe, qui sont les plus grands sous-graphes qui sont connexes.



Au sein d'un graphe connexe, la *distance* entre deux sommets X et Y est définie comme le nombre minimal d'arêtes qu'il faut emprunter pour atteindre Y en partant de X.

Par exemple, la distance entre les sommets 6 et 8 est égale à 3 : le plus court chemin de 6 à 8 est :  $6 \rightarrow 4 \rightarrow 5 \rightarrow 8$ .

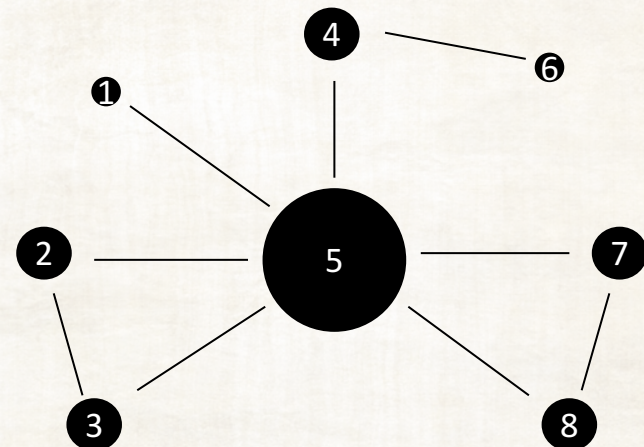
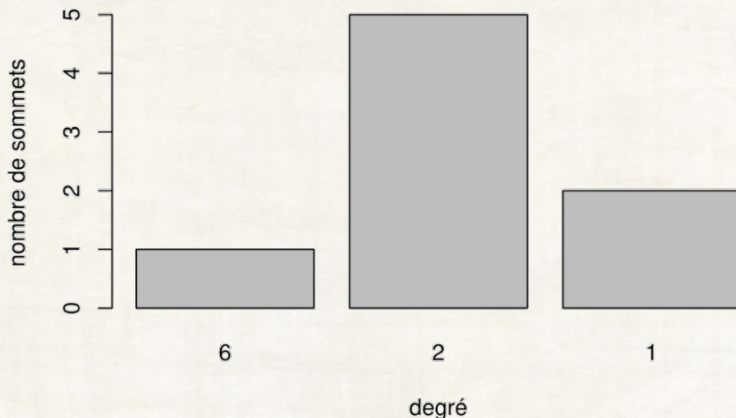
# Degré

On appelle *degré* d'un sommet le nombre total d'arêtes reliant celui-ci aux autres sommets du graphe.

Les individus ayant un degré élevé sont très impliqués dans le réseau correspondant.

Une pratique courante tend à représenter les sommets d'un réseau en utilisant des figures géométriques dont la taille est proportionnelle aux degrés : on peut ainsi en un coup d'œil en identifier les acteurs principaux.

À un niveau plus global, l'analyse de la distribution des degrés de tous les sommets d'un graphe (par exemple à l'aide d'un histogramme) peut être porteuse de beaucoup d'informations.



# Centralité de proximité

La *centralité* d'un sommet est une mesure de son importance au sein du graphe ; le degré peut être considéré comme une telle mesure.

La *centralité de proximité* est définie de la manière suivante : pour un sommet  $s$ , il s'agit de mesurer si  $s$  est « proche » de tous les autres sommets :

$$\text{centralité de proximité } (s) = \frac{1}{\sum_{t \in R} \text{dist}(s, t)}$$

Cette définition est très intuitive :

- si le sommet  $s$  est très « important » au sein du graphe, sa distance au reste du graphe est petite, et sa centralité est donc élevée ;
- au contraire, s'il ne s'agit pas d'un sommet « important », sa distance au reste du graphe est grande et sa centralité est donc petite.

Cette métrique ne signifie pas grand chose par elle-même (il s'agit de l'inverse d'une distance, une variable dont la valeur est difficile à évaluer) : elle n'est vraiment utile que dans une optique de comparaison.

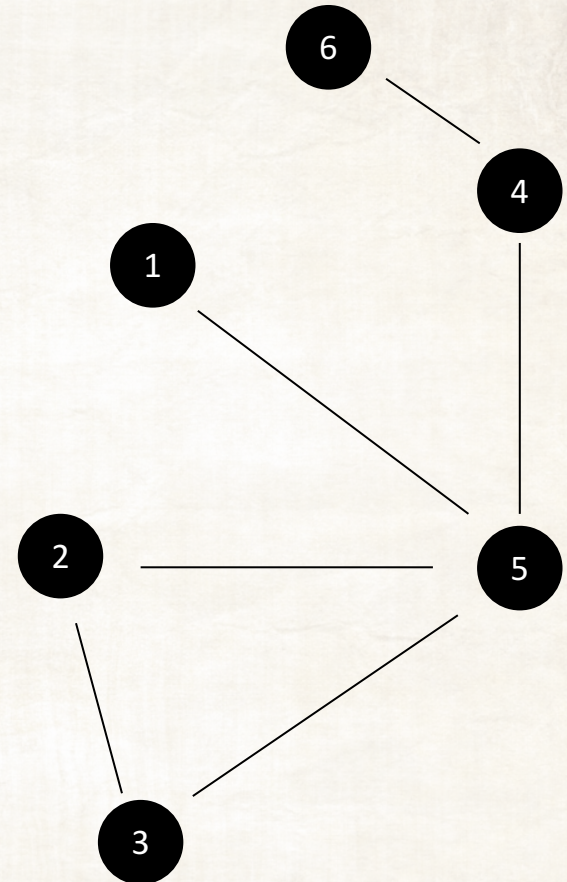
# Centralité de proximité : exemple

centralité (sommets 6)

$$\begin{aligned} &= \frac{1}{\text{dist}(6,1) + \text{dist}(6,2) + \text{dist}(6,3) + \text{dist}(6,4) + \text{dist}(6,5)} \\ &= \frac{1}{3 + 3 + 3 + 1 + 2} = \frac{1}{12} \end{aligned}$$

centralité (sommets 5)

$$\begin{aligned} &= \frac{1}{\text{dist}(5,1) + \text{dist}(5,2) + \text{dist}(5,3) + \text{dist}(5,4) + \text{dist}(5,6)} \\ &= \frac{1}{1 + 1 + 1 + 1 + 2} = \frac{1}{6} \end{aligned}$$



Donc le sommet 5 est plus « central » que le sommet 6, conformément à l'intuition.

# Centralité d'intermédiation

La centralité d'intermédiation (*Betweenness centrality*) d'un sommet  $s$  compte le nombre de plus courts chemins (allant d'un nœud  $i$  à un nœud  $j$ ) en passant par  $s$ . Ici, pour les nœuds 5 et 6 :

Proportion de plus courts chemins passant par 5		1	2	3	4	6
	1		1	1	0	0
	2			0	1	1
	3				1	1
	4					0
	6					

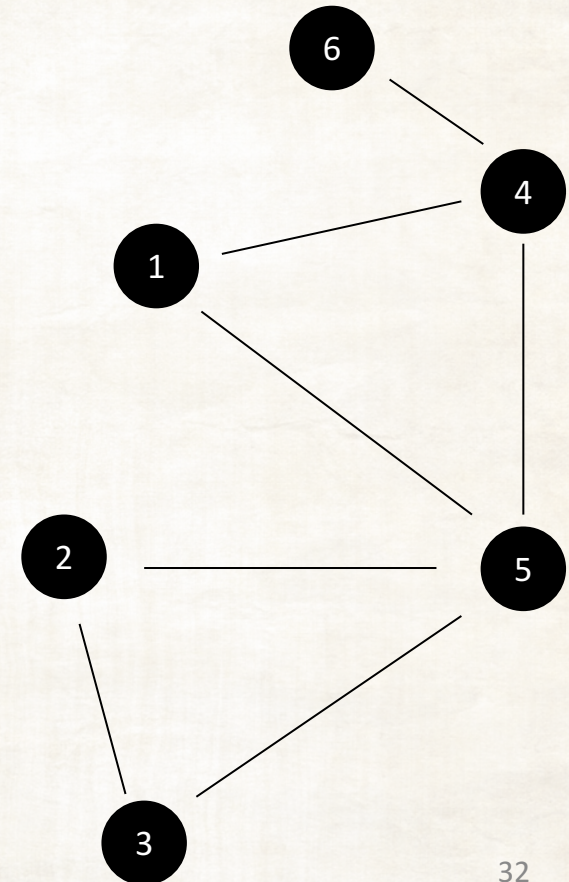
Intermédiation (nœud 5) = 6

(sur un maximum de 10)

Proportion de plus courts chemins passant par 6		1	2	3	4	5
	1		0	0	0	0
	2			0	0	0
	3				0	0
	4					0
	5					

Intermédiation (nœud 6) = 0

(sur un maximum de 10)





# Centralité de vecteur propre

La centralité de vecteur propre (*Eigenvector centrality*) attribue des scores à tous les sommets, de sorte qu'un sommet a un score élevé s'il est connecté à des sommets qui ont eux-mêmes des scores élevés.

Une façon de le calculer est d'appliquer ce processus itératif:

1. dans un premier temps, tous les scores sont égaux à 1
2. le score d'un sommet est égal à la somme des scores de ses voisins :

$$\text{score}_1(2) = \text{score}_0(3) + \text{score}_0(5)$$

$$\text{score}_1(4) = \text{score}_0(1) + \text{score}_0(5) + \text{score}_0(6)$$

...

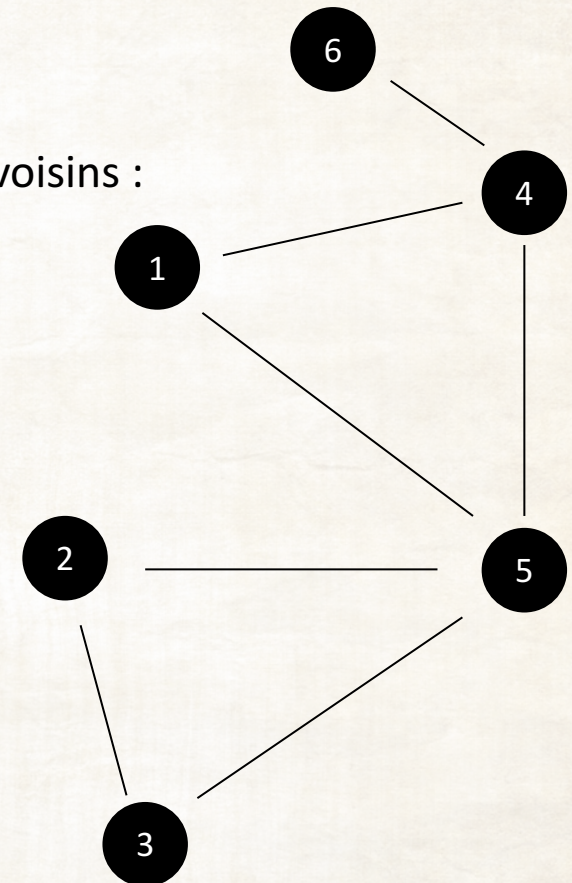
$$\text{score}_2(2) = \text{score}_1(3) + \text{score}_1(5)$$

$$\text{score}_2(5) = \text{score}_1(1) + \text{score}_1(2) + \text{score}_1(3) + \text{score}_1(4)$$

$$\text{score}_2(6) = \text{score}_1(4)$$

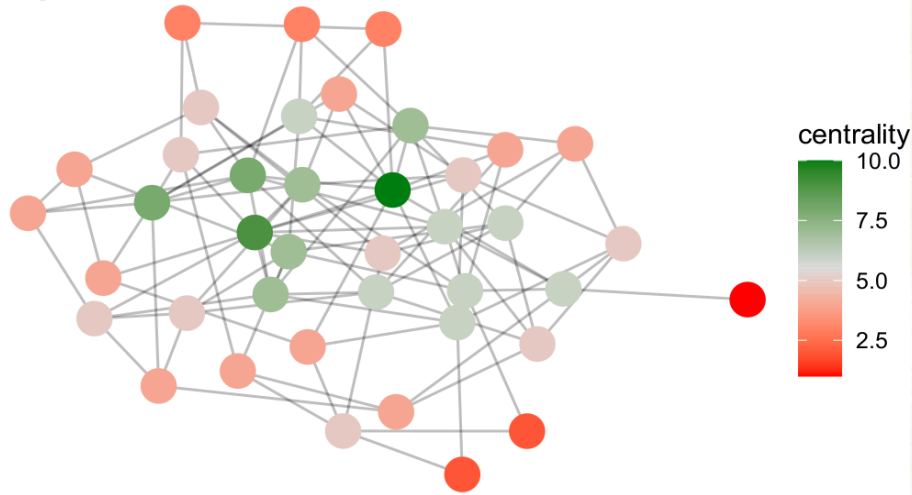
3. après 100 itérations, on normalise les scores obtenus pour obtenir la centralité de vecteur propre :

1	2	3	4	5	6
17%	16%	16%	19%	25%	7%

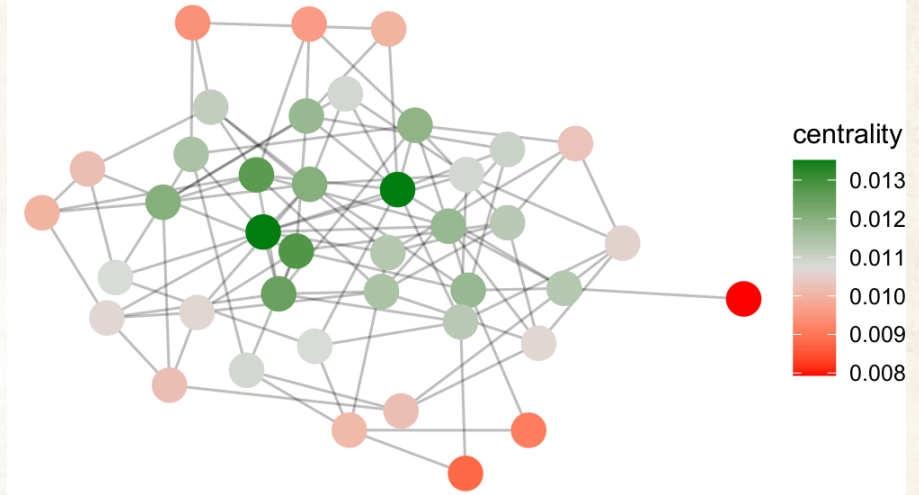


# Différentes métriques, différentes utilisations

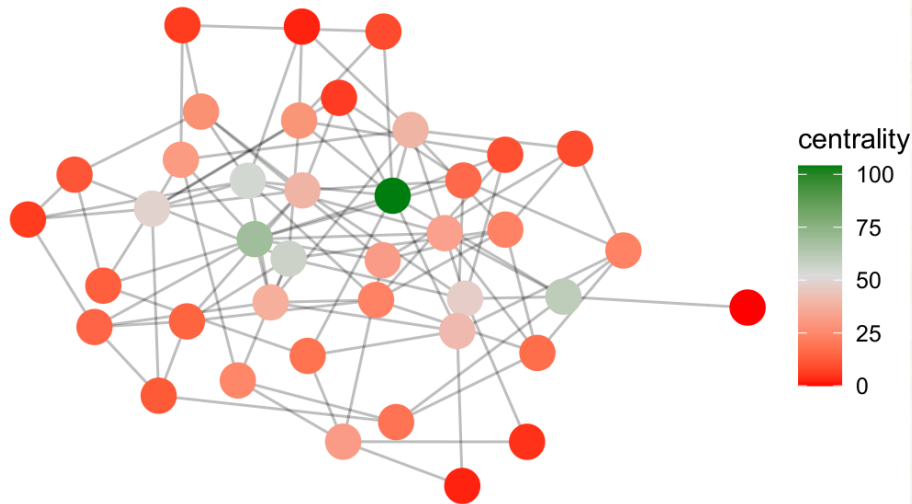
Degree



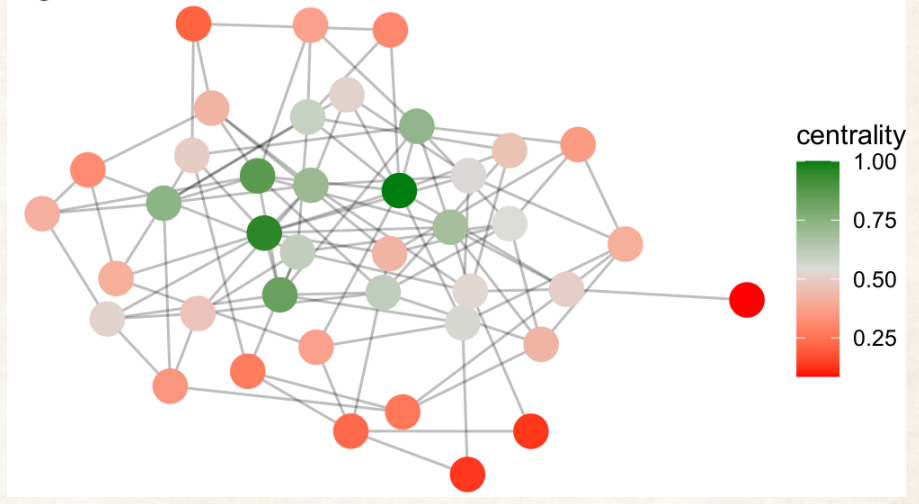
Closeness



Betweenness



Eigenvector



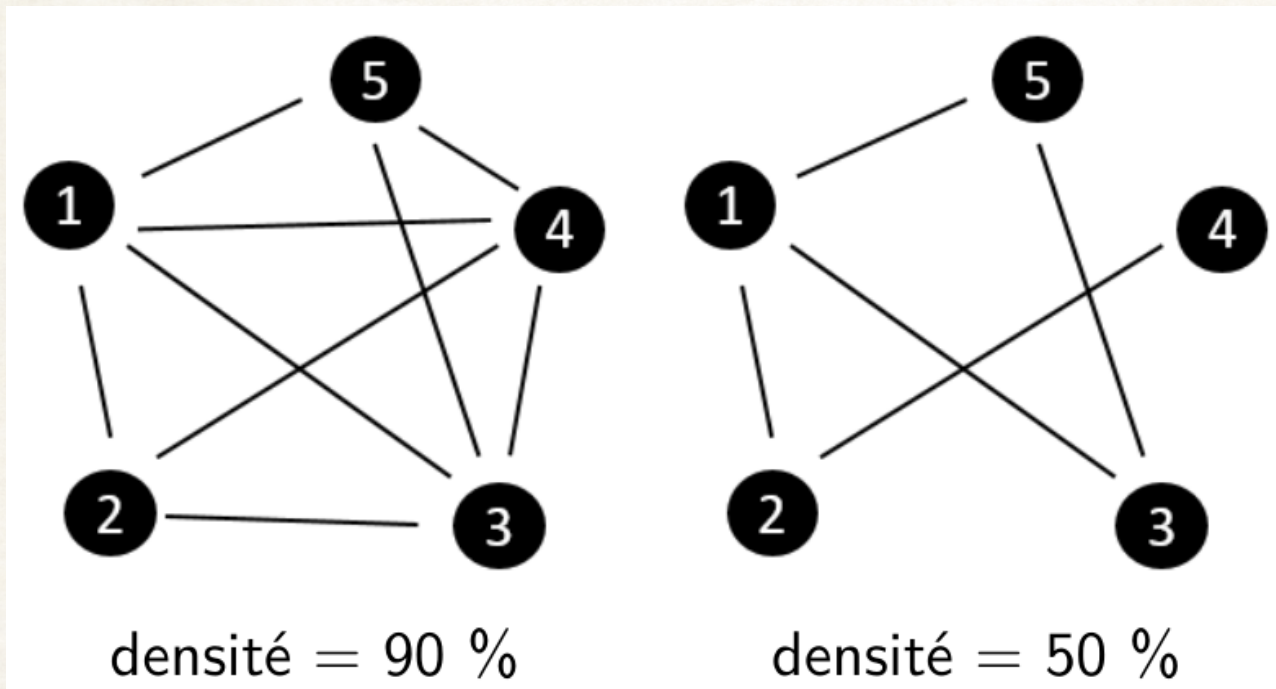
## Autres notions de centralité

Sans entrer dans les détails, mentionnons quelques autres mesures de centralité :

- *centralité d'intermédiarité* : mesure du nombre de plus courts chemins sur lesquels le sommet est situé ;
- *centralité spectrale* : mesure qui attribue itérativement des scores aux sommets en se basant sur les scores de leurs proches voisins ;
- *centralité d'information* ;
- *centralité de Katz* ;
- *centralité de percolation* ;
- ...

# La notion de densité

- La densité d'un réseau est la proportion d'arêtes dans le réseau par rapport au nombre maximal d'arêtes qu'un réseau ayant le même nombre de sommets peut posséder.



# Homophilie

L'homophilie d'un réseau est la tendance des individus qui le composent à s'associer avec d'autres individus qui leur sont semblables selon un critère choisi.

Le coefficient d'homophilie est défini pour un graphe non orienté  $R$  et un groupe de  $C$  catégories (permettant de classer chacun des sommets de  $R$ ) de la manière suivante :

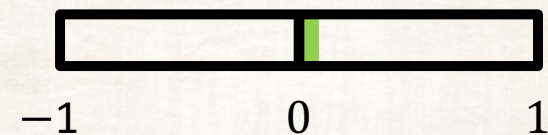
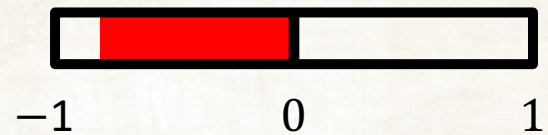
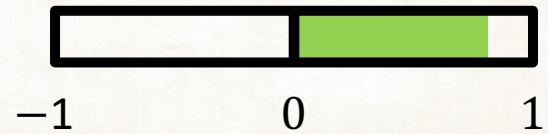
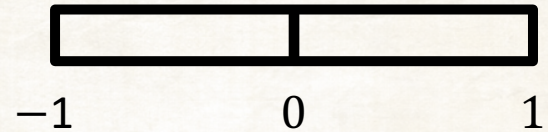
$$\text{coefficient d'homophilie } (R) = \frac{\sum_{i=1}^C p_{ii} - \sum_{i=1}^C \left( \sum_{j=1}^C p_{ij} \right)^2}{1 - \sum_{i=1}^C \left( \sum_{j=1}^C p_{ij} \right)^2}$$

où  $p_{ij}$  est la portion des arêtes de  $R$  qui relie un sommet appartenant à la catégorie  $i$  à un sommet de la catégorie  $j$ .

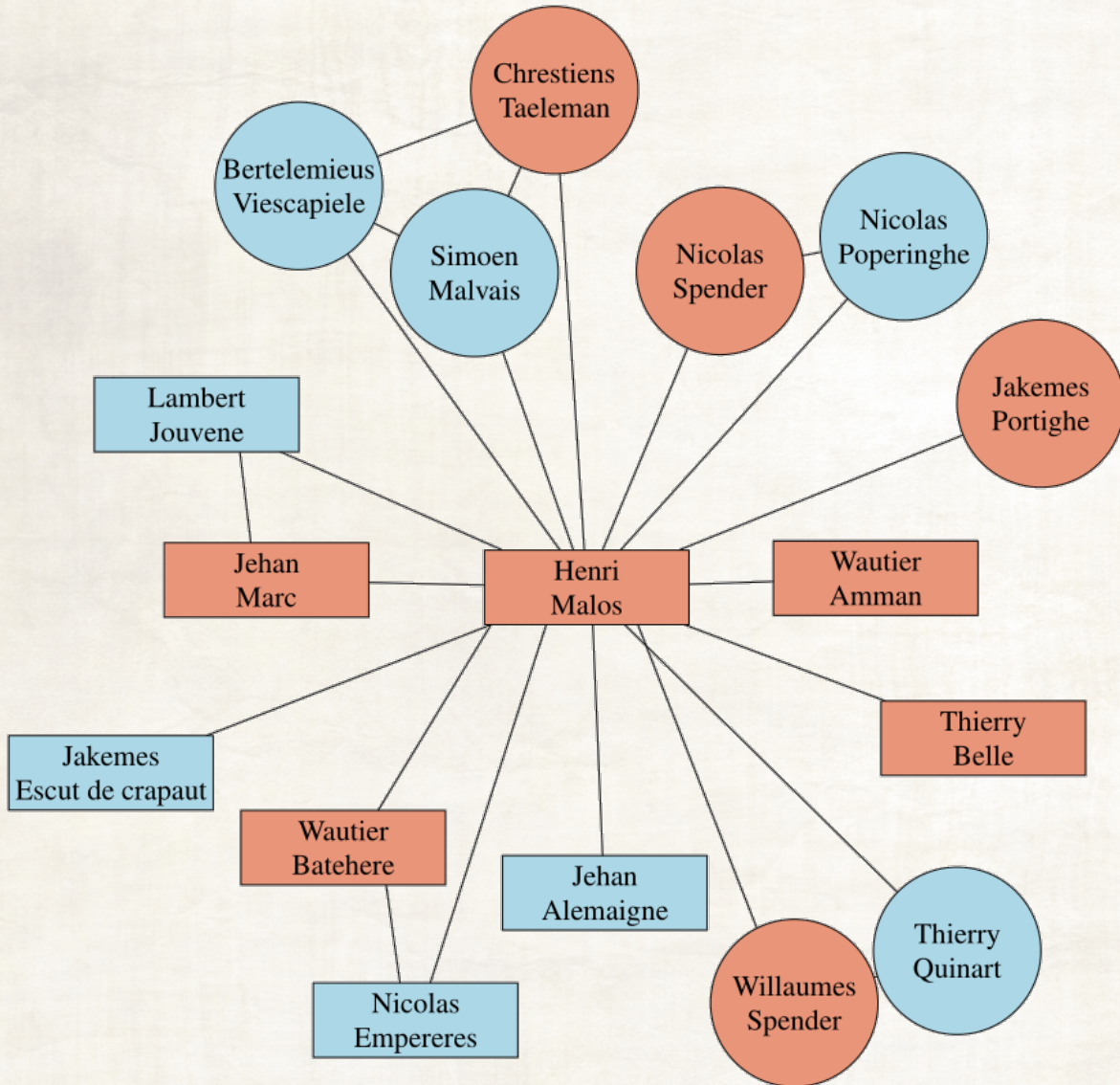
# Homophilie : intuition

On peut voir cette métrique comme une mesure de « corrélation » des catégories avec elles-mêmes :

- le coefficient d'homophilie est toujours compris entre  $-1$  et  $1$  ;
- lorsqu'il est **positif**, le réseau est *homophile*, et ses sommets ont tendance à être connectés avec d'autres sommets de la même catégorie ;
- au contraire, lorsqu'il est **négatif**, le réseau est *dishomophile*, et ses sommets ont tendance à être connectés avec des sommets d'autres catégories ;
- lorsqu'il est proche de **zéro**, le réseau est *anhomophile*, et les catégories n'exercent pas d'influence sur les relations entre les sommets.



# Homophilie : exemple



Le coefficient d'homophilie de ce réseau pour les catégories de formes est positif, avec une valeur de 31,85%.

→ les individus-cercles sont plutôt reliés à d'autres individus-cercles, les individus-rectangles sont plutôt reliés à d'autres individus-rectangles.

Le coefficient d'homophilie de ce réseau pour les catégories de couleurs est négatif, avec une valeur de -11,03%.

→ les individus bleus sont plutôt reliés à de individus rouges, et vice-versa.

# La détection de communautés

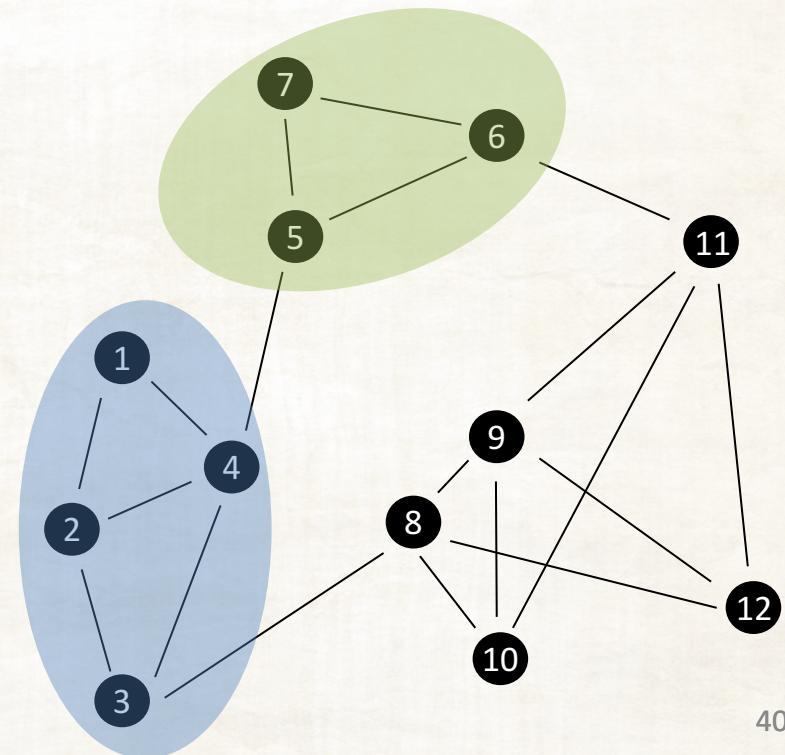
Une communauté est un ensemble d'entités qui interagissent plus souvent entre elles qu'avec d'autres. Graphiquement, une communauté est un groupe de sommets fortement liés les uns aux autres et faiblement liés à d'autres sommets.

Il existe de nombreuses méthodes différentes pour détecter automatiquement les communautés.

L'algorithme *min-cut* est l'une de ces méthodes.

Il vise à diviser un graphe connecté en un nombre prédéterminé de sous-graphes connectés disjoints en supprimant le moins d'arêtes possible.

Les groupes de sommets formant les sous-graphes ainsi obtenus sont les communautés.

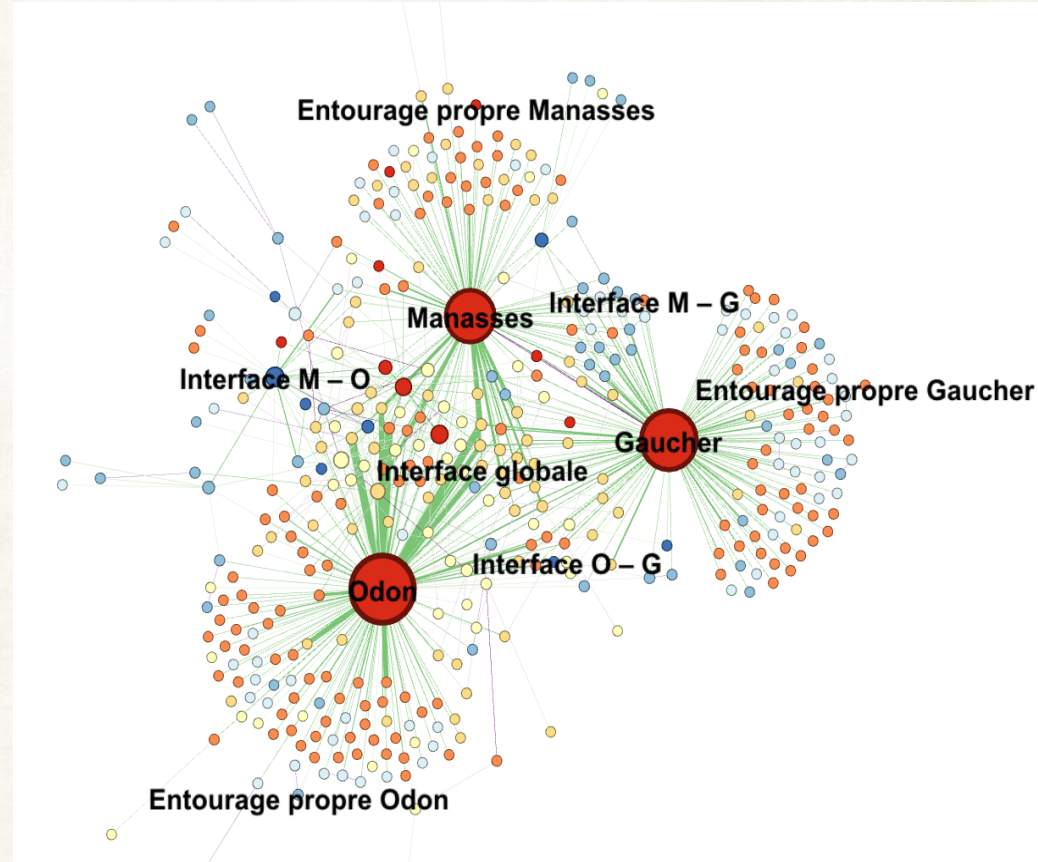
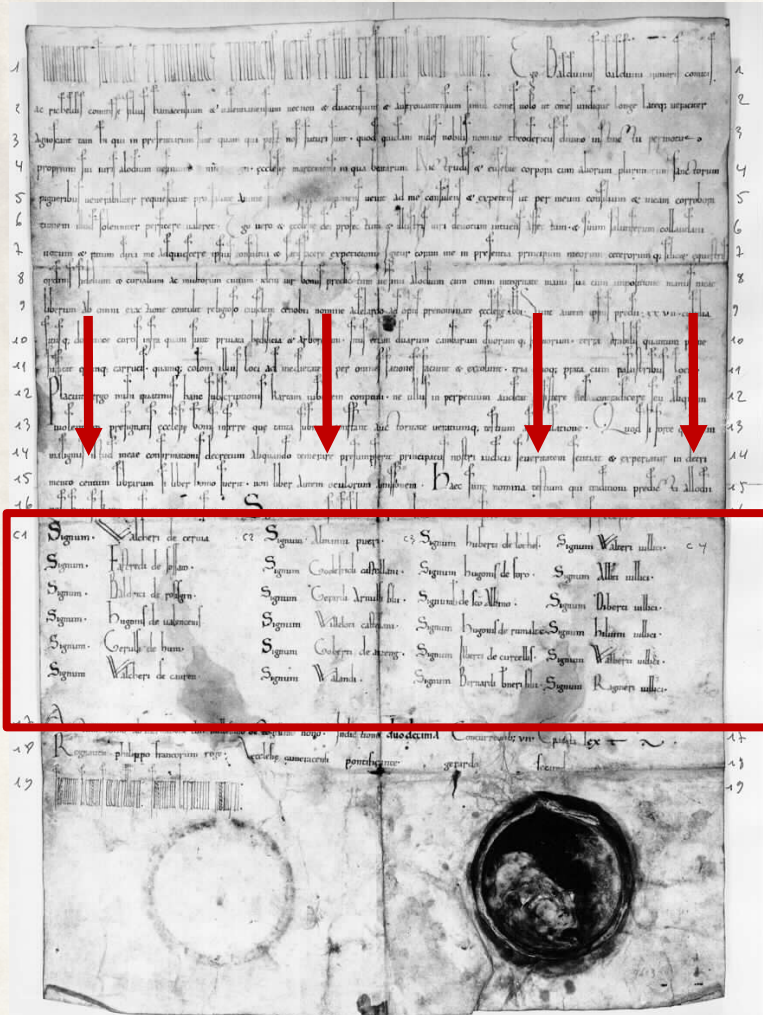




# Quelques exemples d'utilisation

Une approche « classique » : développer une analyse politique au départ de chartes

## Exemple : la Querelle des Investitures à Cambrai (1092-1113)

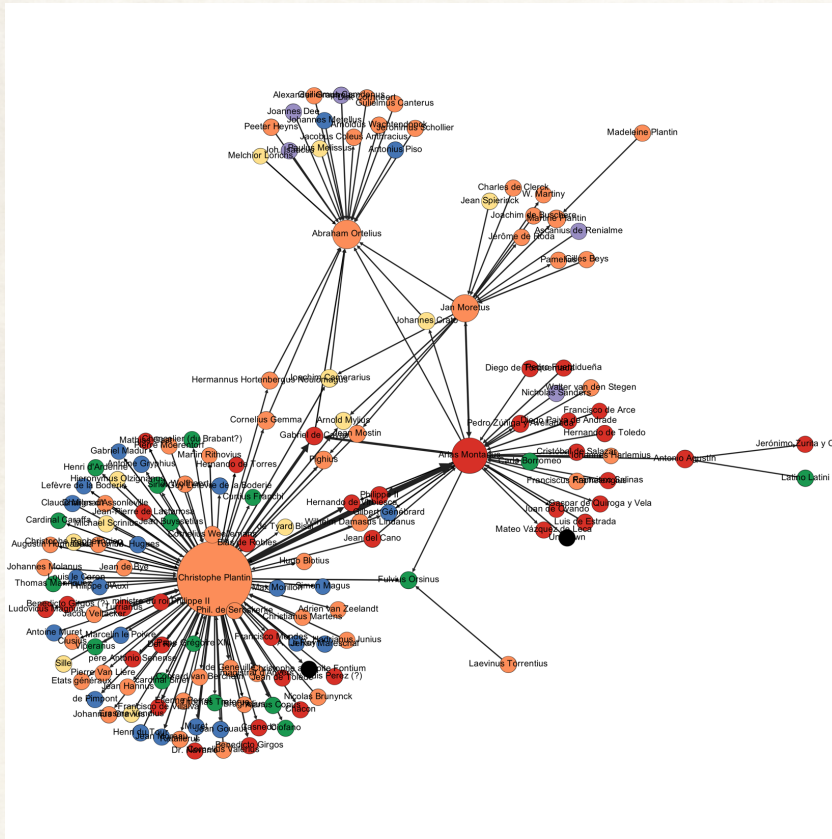


Source : Nicolas Ruffini-Ronzani, « Relire l'histoire des principautés territoriales à travers l'analyse de réseaux... »

# Quelques exemples d'utilisation

Une autre approche « classique » : l'analyse des réseaux épistolaires

## Exemple : le réseau épistolaire de l'imprimeur Christophe Plantin (1520-1589)



### Légende

Orange = anciens Pays-Bas

Rouge = péninsule Ibérique

Vert = Italie

Bleu = France

Jaune = Empire

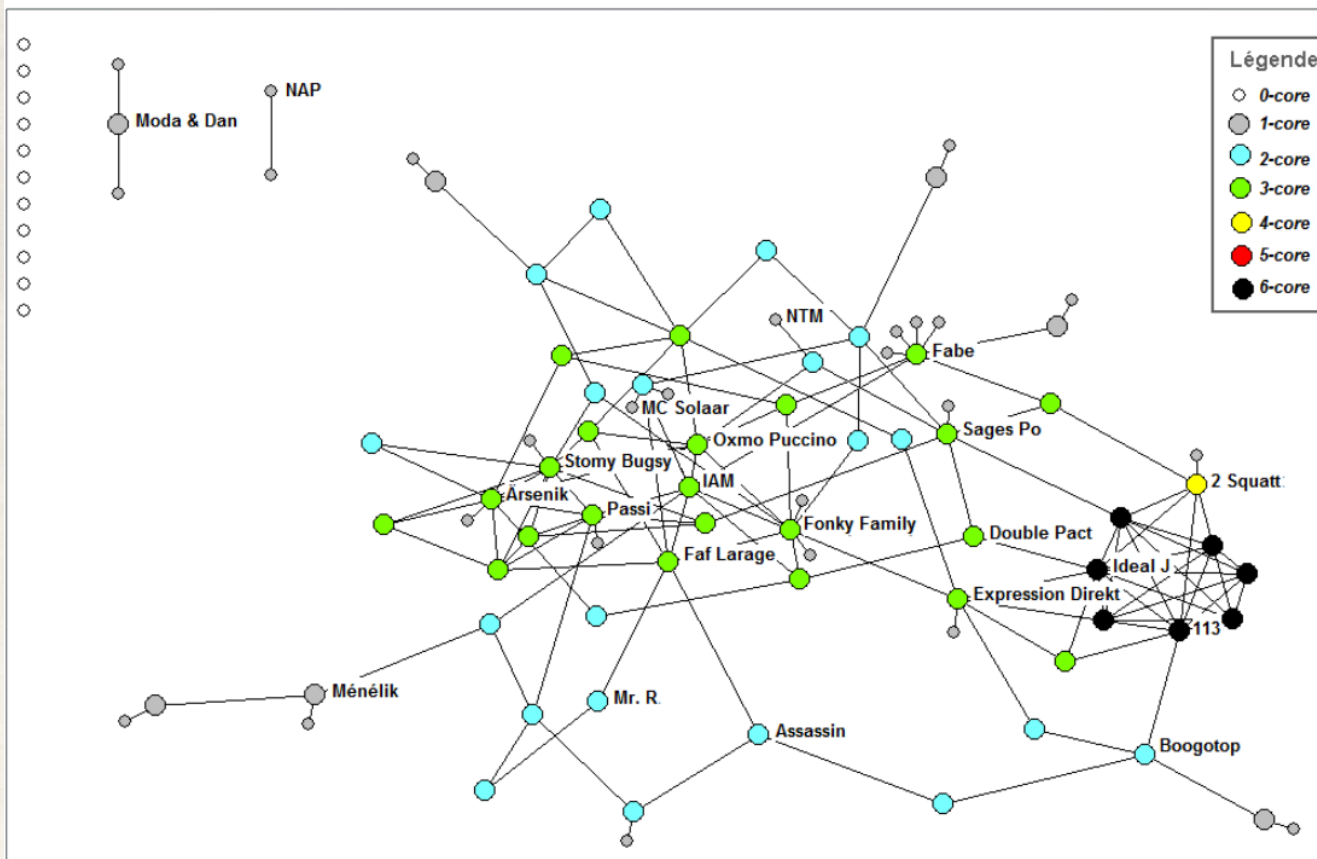
Bleu/mauve = Angleterre

Noir = Inconnu

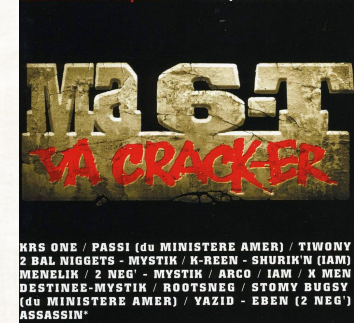
Source : exercice préparé dans le cadre du cours de Valérie Leyh, « Le livre et la culture numérique », 2020-2021 (UNamur).

# Quelques exemples d'utilisation

Analyse des collaborations dans le milieu culturel  
Exemple : le rap français dans les années 1990-2000



du film de Jean-François RICHEL Musique de WHITE & SPIRIT



Source : Karim Hammou, « Des raps en français au 'rap français'.. », *Histoire et mesure*, 24 (2009), p. 73-108. Accès en ligne : <https://www.cairn.info/revue-histoire-et-mesure-2009-1-page-73.htm>

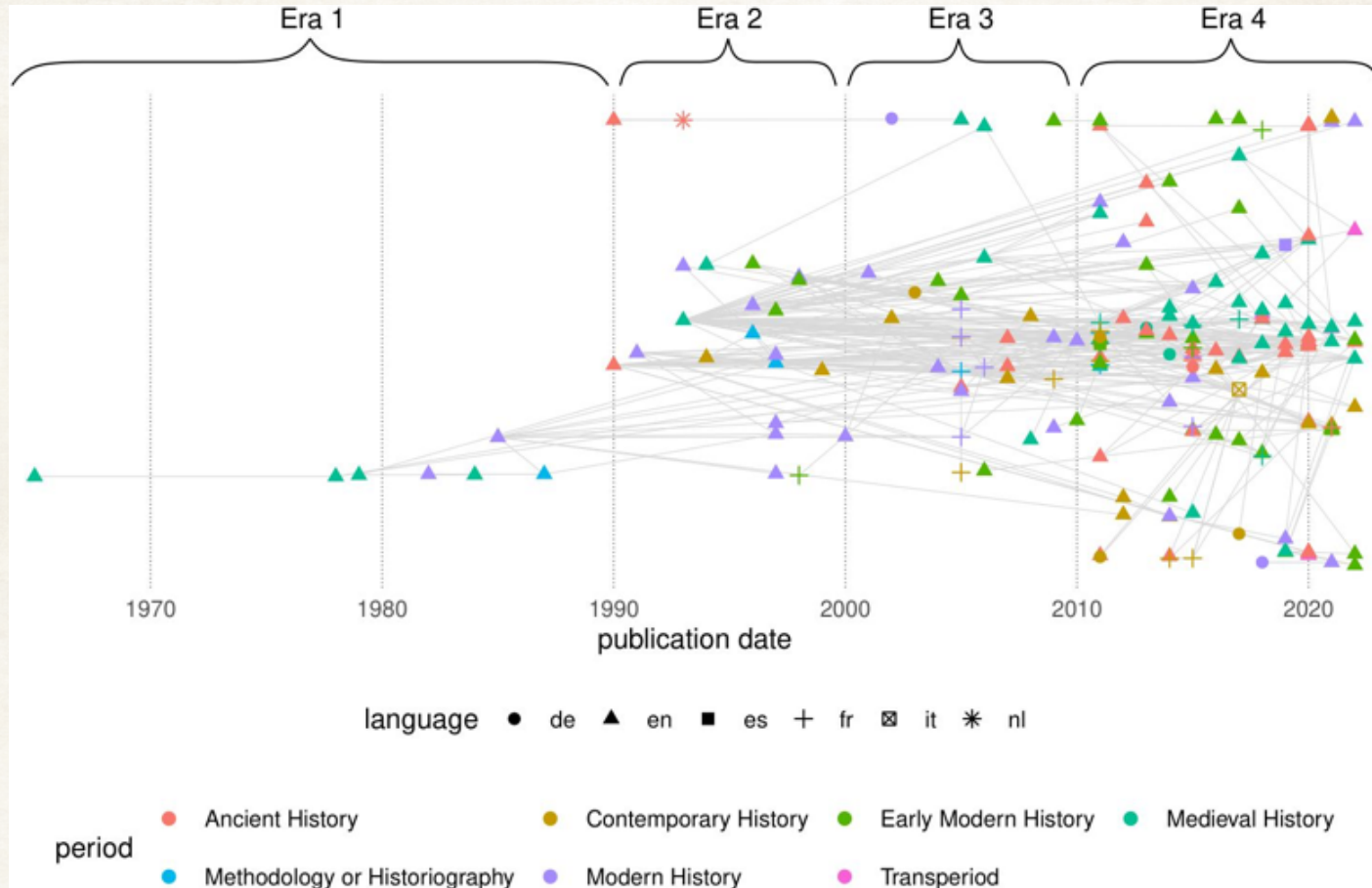




# Quelques exemples d'utilisation

Analyser un réseau de citations

## Exemple : l'utilisation de l'analyse de réseaux en Histoire



**Source :** Nicolas Ruffini-Ronzani, Sébastien de Valeriola, « A Network History of Historical Network Analysis. Using a Citation Network to Explore Historiography », soumis.