

Cours *Distant Reading* : Visualisation

7. Introduction à l'analyse de réseau

Simon Gabay



Introduction

SOLVTIO PROBLEMATIS

AD

GEOMETRIAM SITVS

PERTINENTIS.

AVCTORE

Leont. Euler.

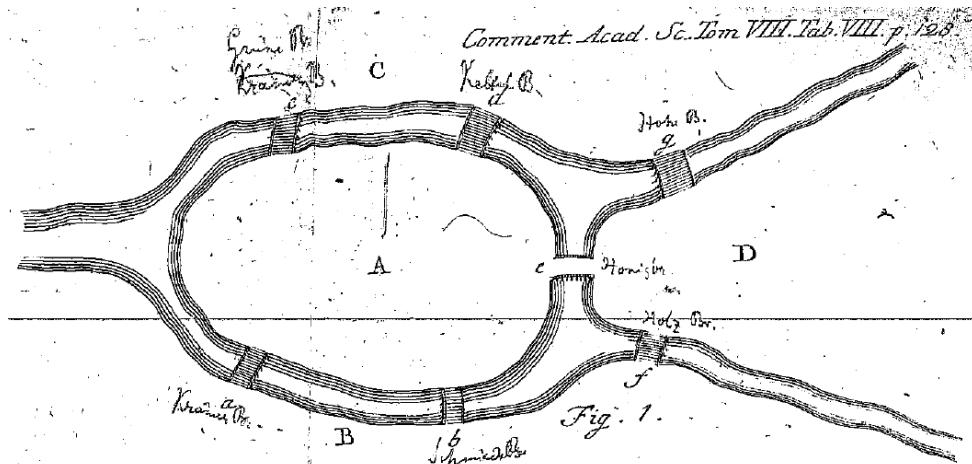
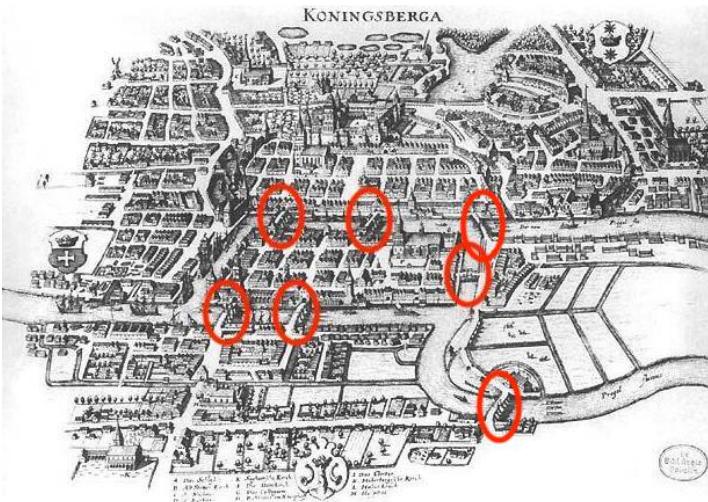
§. I.

Tabula VIII. Praeter illam Geometriae partem, quae circa quantitates versatur, et omni tempore summo studio est exculta, alterius partis etiamnum admodum ignotae primus mentionem fecit *Leibnitzius*, quam Geometriam situs vocavit. Ista pars ab ipso in solo situ determinando, situsque proprietatibus eruendis occupata esse statuitur; in quo negotio neque ad quantitates respiciendum, neque calculo quantitatum vtendum sit. Cuiusmodi autem problemata ad hanc situs Geometriam pertineant, et quali methodo in iis resoluendis vti oporteat, non satis est definitum. Quamobrem, cum nuper problematis cuiusdam mentio esset facta, quod quidem ad geometriam pertinere videbatur, at ita erat comparatum, vt neque determinationem quantitatum requireret, neque solutionem calculi quantitatum ope admitteret, id ad geometriam situs referre haud dubitauit: praeferentim quod in eius solutione solus situs in considerationem veniat, calculus vero nullius prorsus sit vius. Methodum ergo meam quam ad huius generis proble-

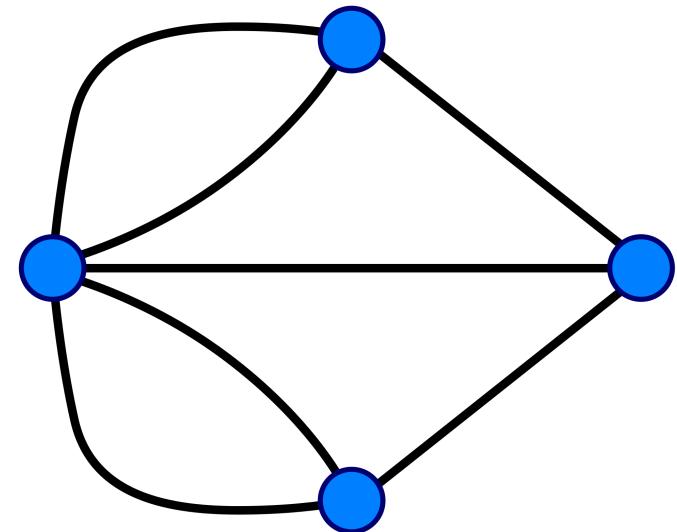
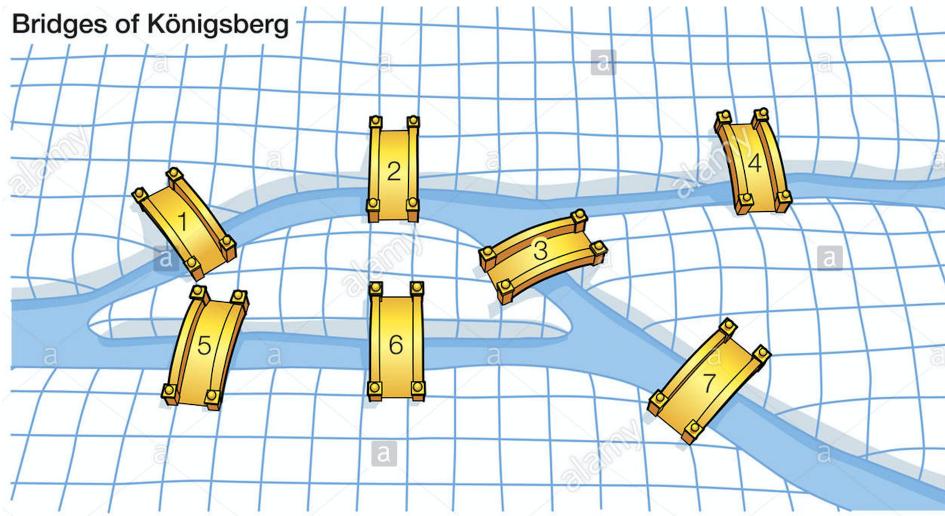
mata

Les sept ponts de Königsberg

- Leonhard Euler (1707-1783), mathématicien et physicien suisse
- «Solutio problematis ad geometriam situs pertinentis», *Mémoires de l'Académie des sciences de Berlin*, 1735
- Est-ce qu'il existe une promenade dans Königsberg permettant, peu importe le point de départ, de revenir à celui-ci en prenant tous les ponts une seule fois?
- Euler est à l'origine de la théorie des graphes parce qu'il fut le premier à proposer un traitement mathématique de la question, suivi par Alexandre-Théophile Vandermonde.



Bridges of Königsberg



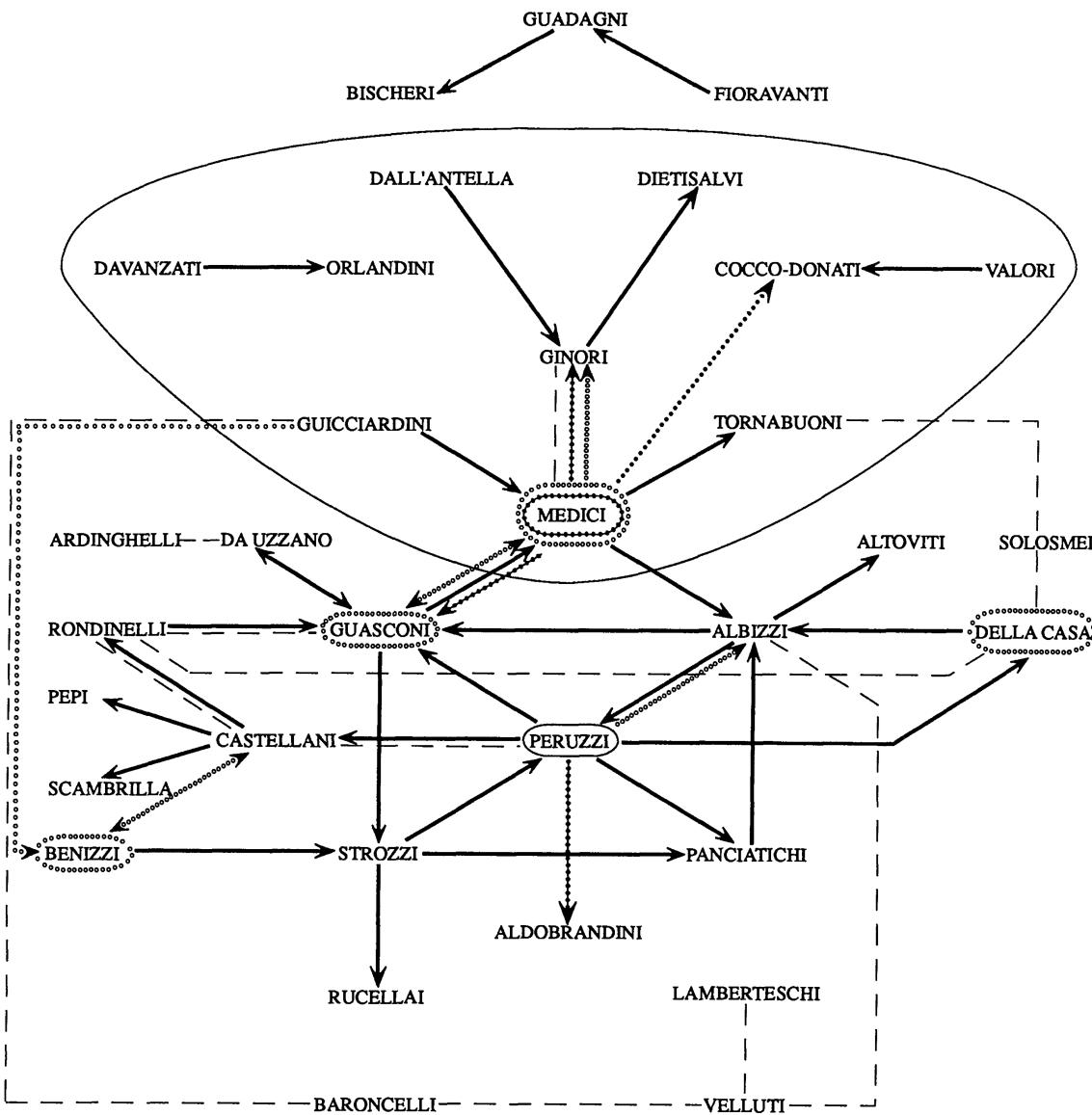
Source: [wikipedia](#), [wikipedia](#), [wikipedia](#), [wikipedia](#)

The Rise of the Medici

to understand state formation one must penetrate beneath the veneer of formal institutions, groups, and goals down to the relational substrata of peoples' actual lives

John F. Padgett et Christopher K. Ansell, Robust Action and the Rise of the Medici, *American Journal of Sociology*, 98(6), 1993

Tentative de représentation graphique des réseaux de sociabilité dans la Florence du XVème s.



Types of Ties:

→ Marriage

..... Partnership

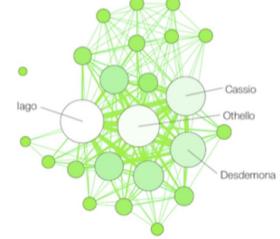
----- Bank Employment

----- Trade

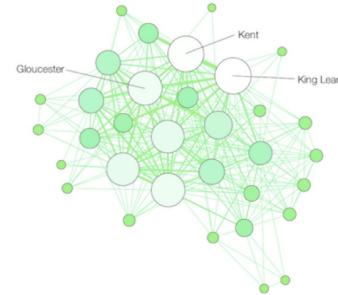
..... Real Estate

FIG. 2a.—Marriage and economic blockmodel structure (92 elite families)

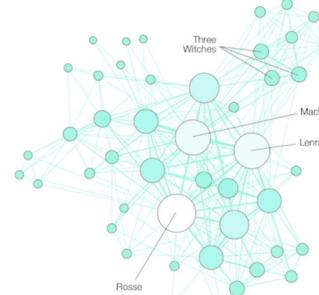
Littérature: Shakespeare



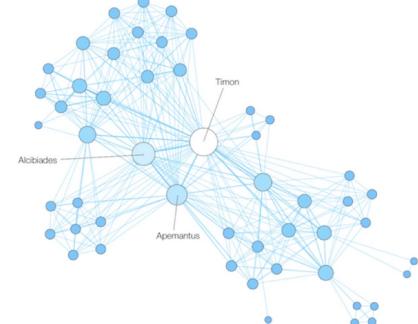
OTHELLO
Number of characters **24** | 55% Network density



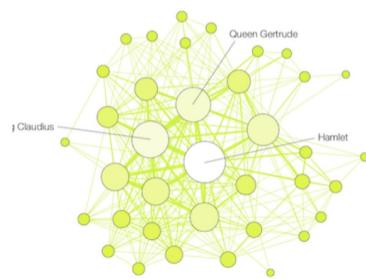
KING LEAR
Number of characters **33** | 45% Network density



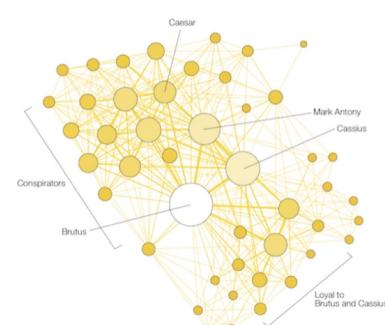
MACBETH
Number of characters **46** | 25% Network density



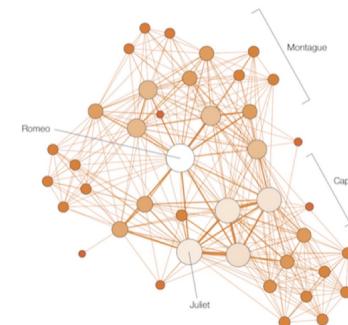
TIMON OF ATHENS
Number of characters **51** | 25% Network density



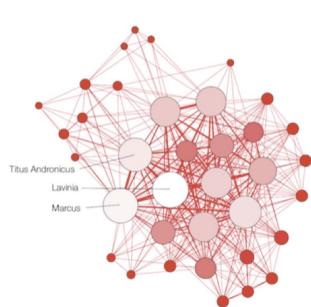
HAMLET
Number of characters **37** | 39% Network density



JULIUS CAESAR
Number of characters **46** | 34% Network density



ROMEO AND JULIET
Number of characters **41** | 37% Network density

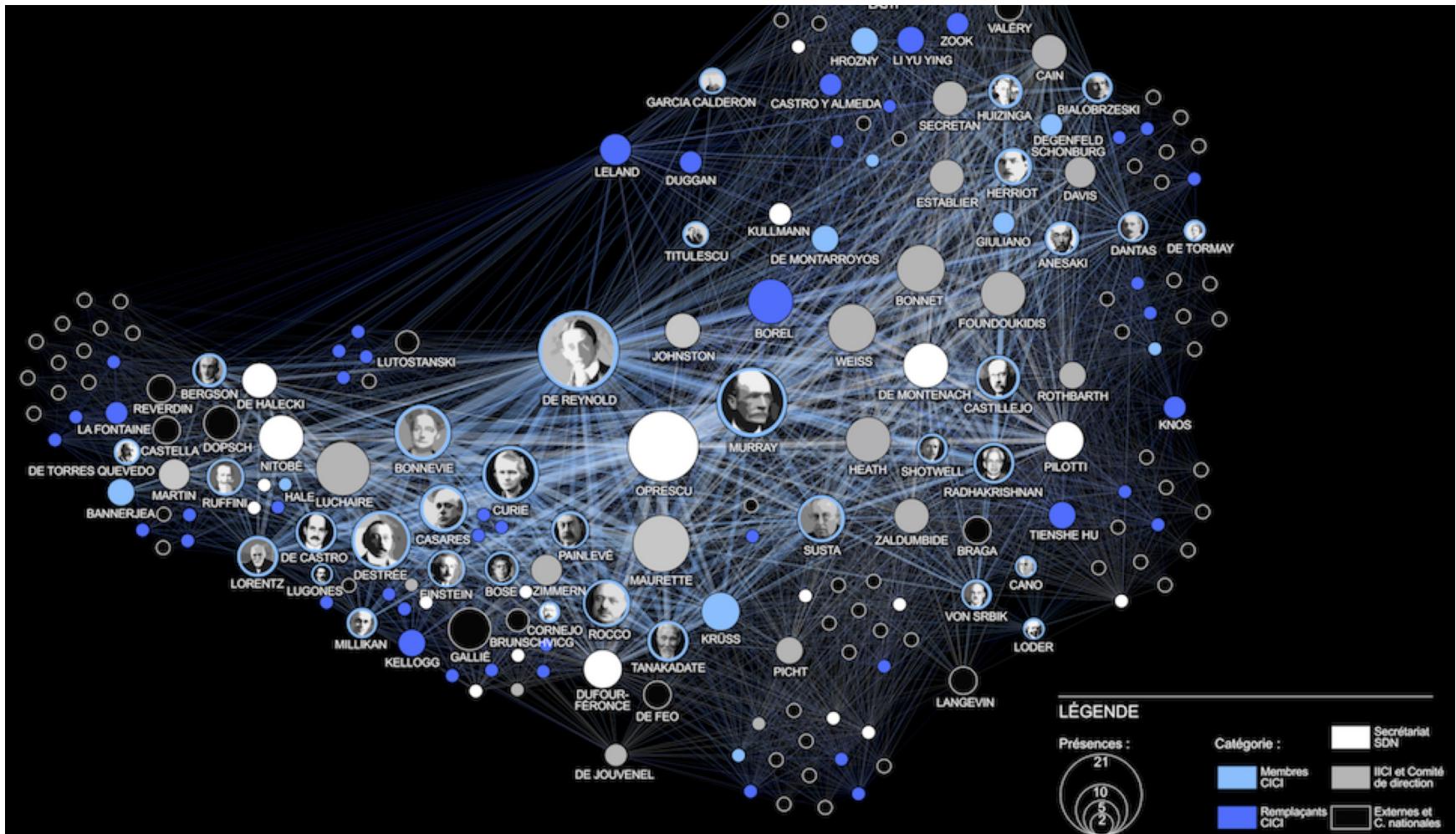


TITUS ANDRONICUS
Number of characters **36** | 50% Network density

M. Grandjean. Network visualization: mapping Shakespeare's tragedies.

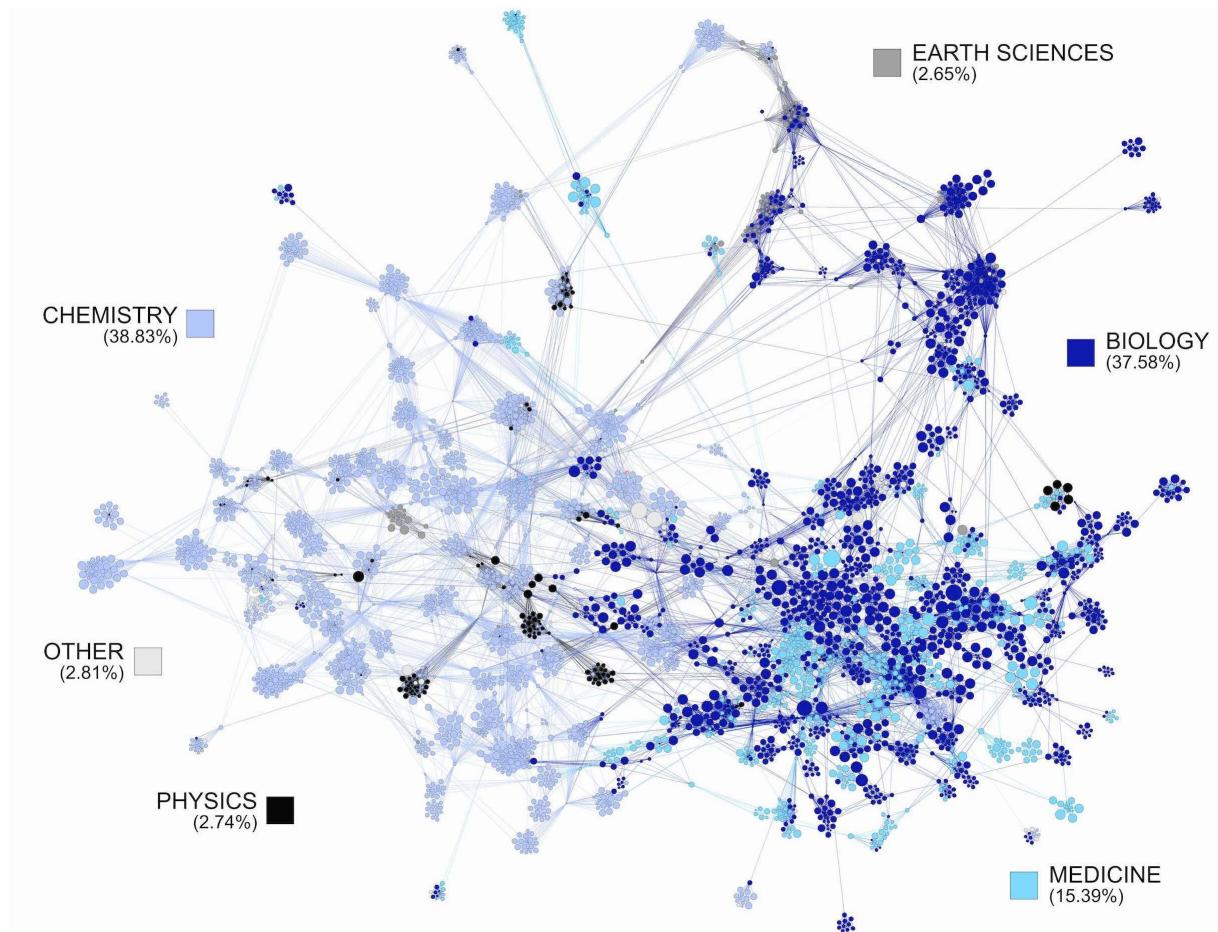
2015 [lien](#)

Histoire: ligue des Nations



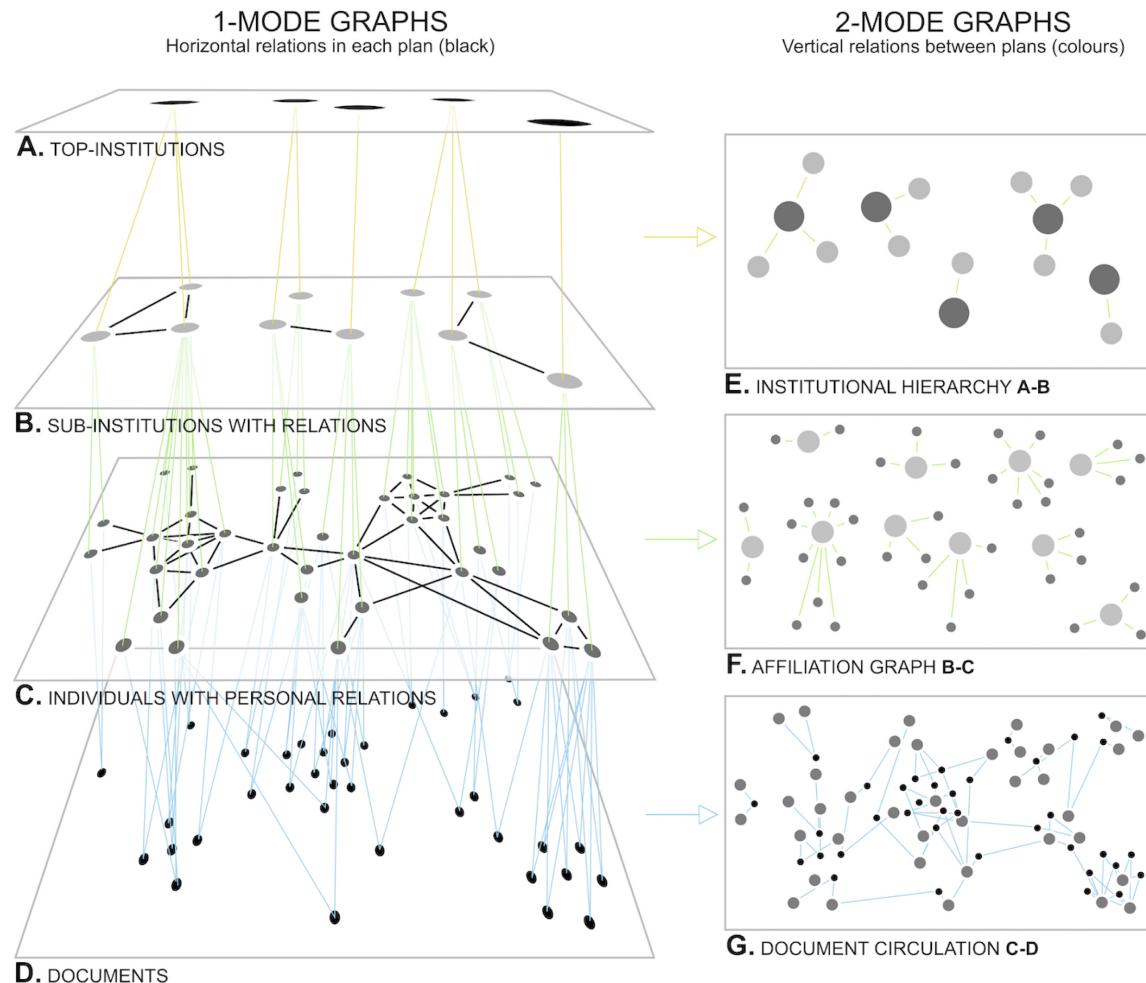
M. Grandjean. Social Network Analysis of the League of Nations' Intellectual Cooperation, an Historical Distant Reading. *DH Benelux*, 2016, Luxembourg. [lien](#)

Sociologie: collaborations entre les chercheurs dans les projets FNS



Cf. M. Grandjean, et al. Complex Network Visualisation for the History of Interdisciplinarity: Mapping Research Funding in Switzerland. *Digital Humanities 2017*, 2017, Montreal, Canada. [lien](#)

Réseaux à plusieurs niveaux



M. Grandjean. Multimode and Multilevel: Vertical Dimension in Historical and Literary Networks. *Digital Humanities 2017*, Aug 2017, Montreal, Canada. [lien](#)

Introduction

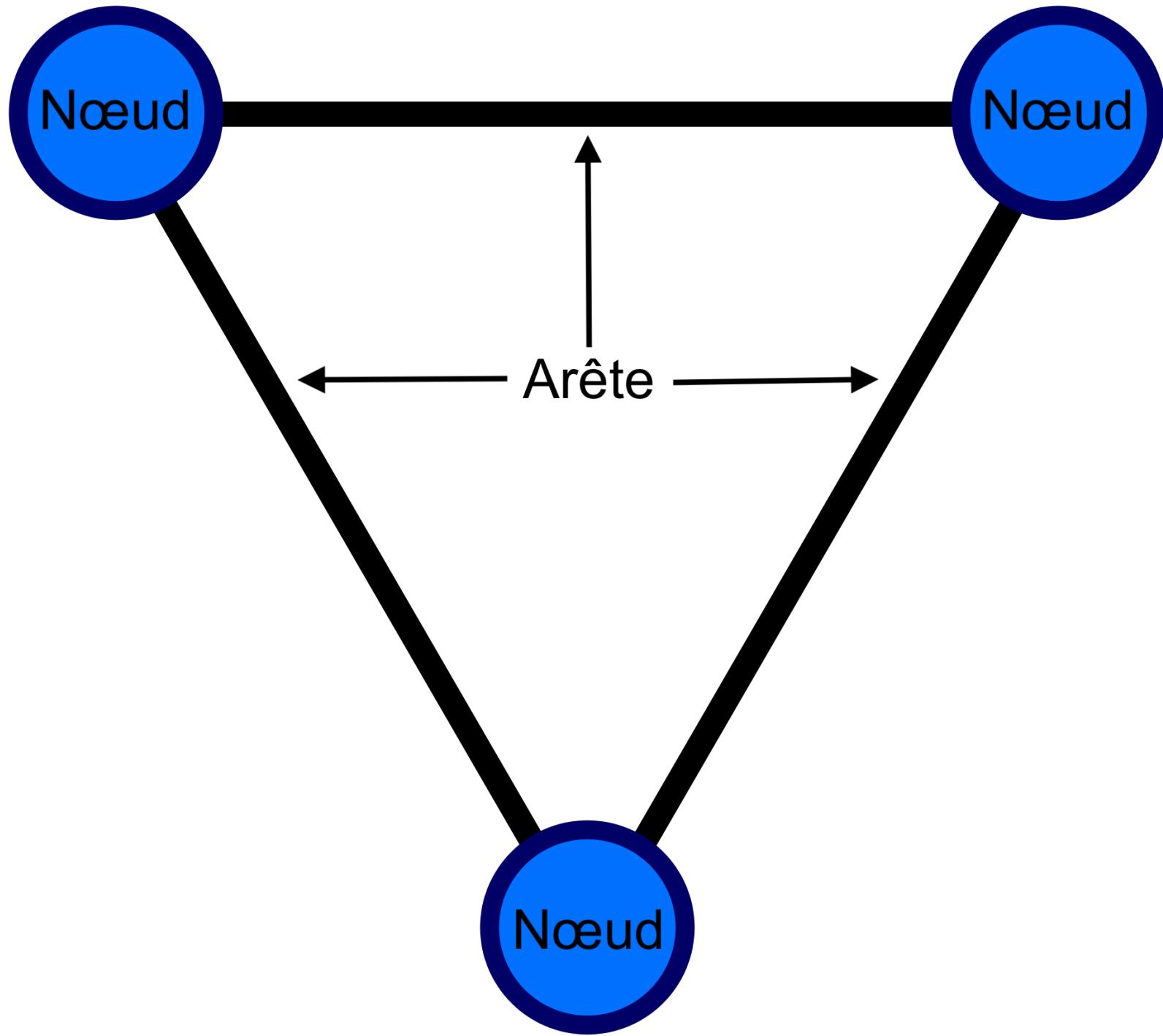
Théorie des graphes

- Comment représenter des relations complexes?
- Le graphe est un modèle abstrait de dessin de réseaux reliant des objets
- C'est aussi une représentation mathématique. Ainsi, dans sa version la plus simple (graphe simple non orienté), un graphe est un couple $G = (V, E)$ où
 - V est un ensemble de sommets
 - E un ensemble d'arêtes, sachant que $E \subseteq \{\{x, y\} / (x, y) \in V^2\}$ c'est-à-dire que E inclus au sens large l'arête (ou paire de sommets) $\{x, y\}$ tel que x, y appartiennent à V au carré.

Nous reviendrons à cette question mathématique plus tard (rapidement).

Définitions

- Un *graphe* est une somme de points et de lignes
- Un point dans le graphe est appelé *nœud* (*node*)
- Une ligne dans le graphe est appelée *arête* ou *arc* (*edge*)
- Un *réseau* est un graphe avec des attributs

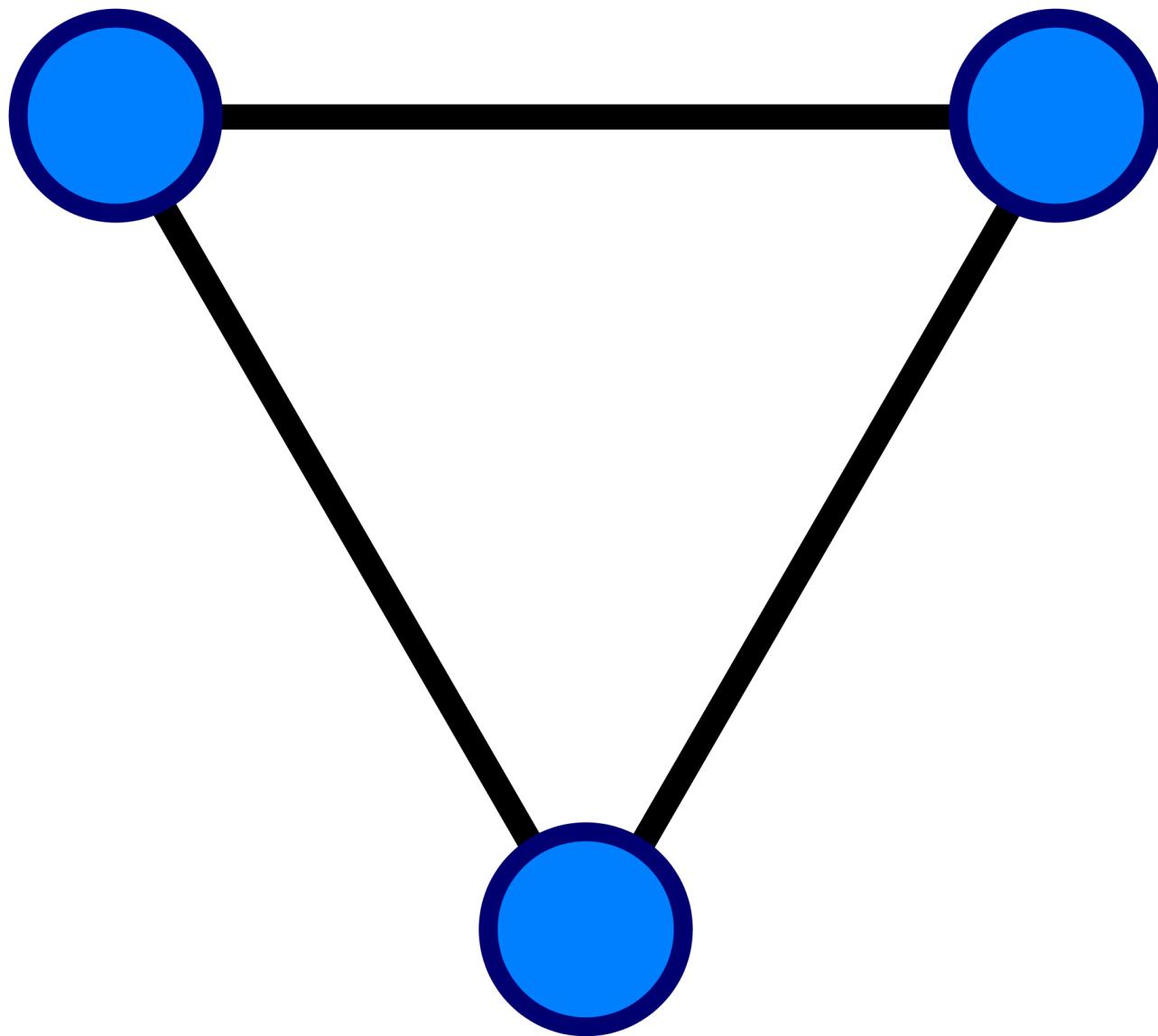


Les types de réseaux

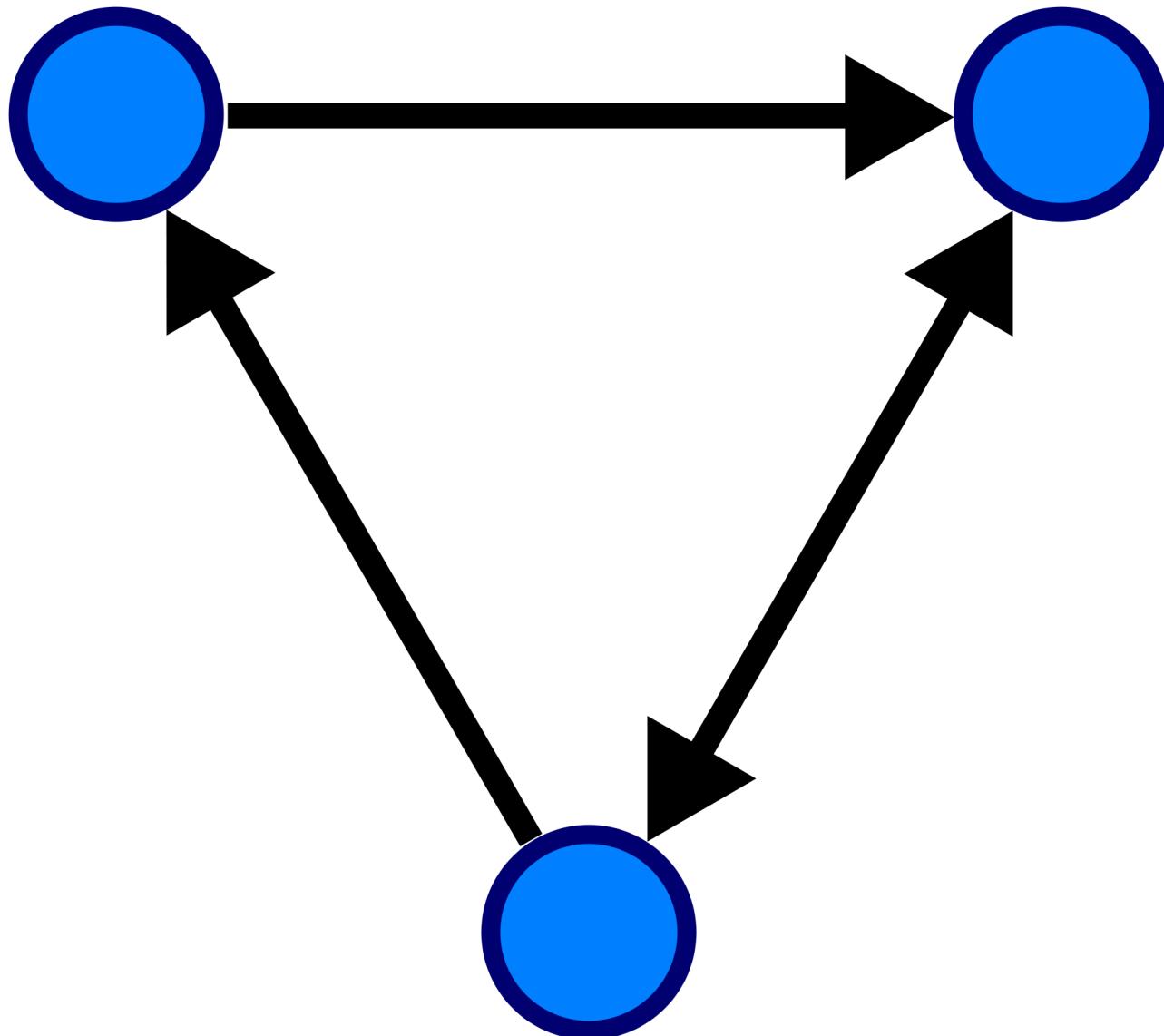
Un réseau peut être

- non-orienté (*undirected graph*) . . .
- . . . orienté (*directed graph*) . . .
- . . . signé (*signed graph*) . . .
- . . . ou valué (*valued graph*)

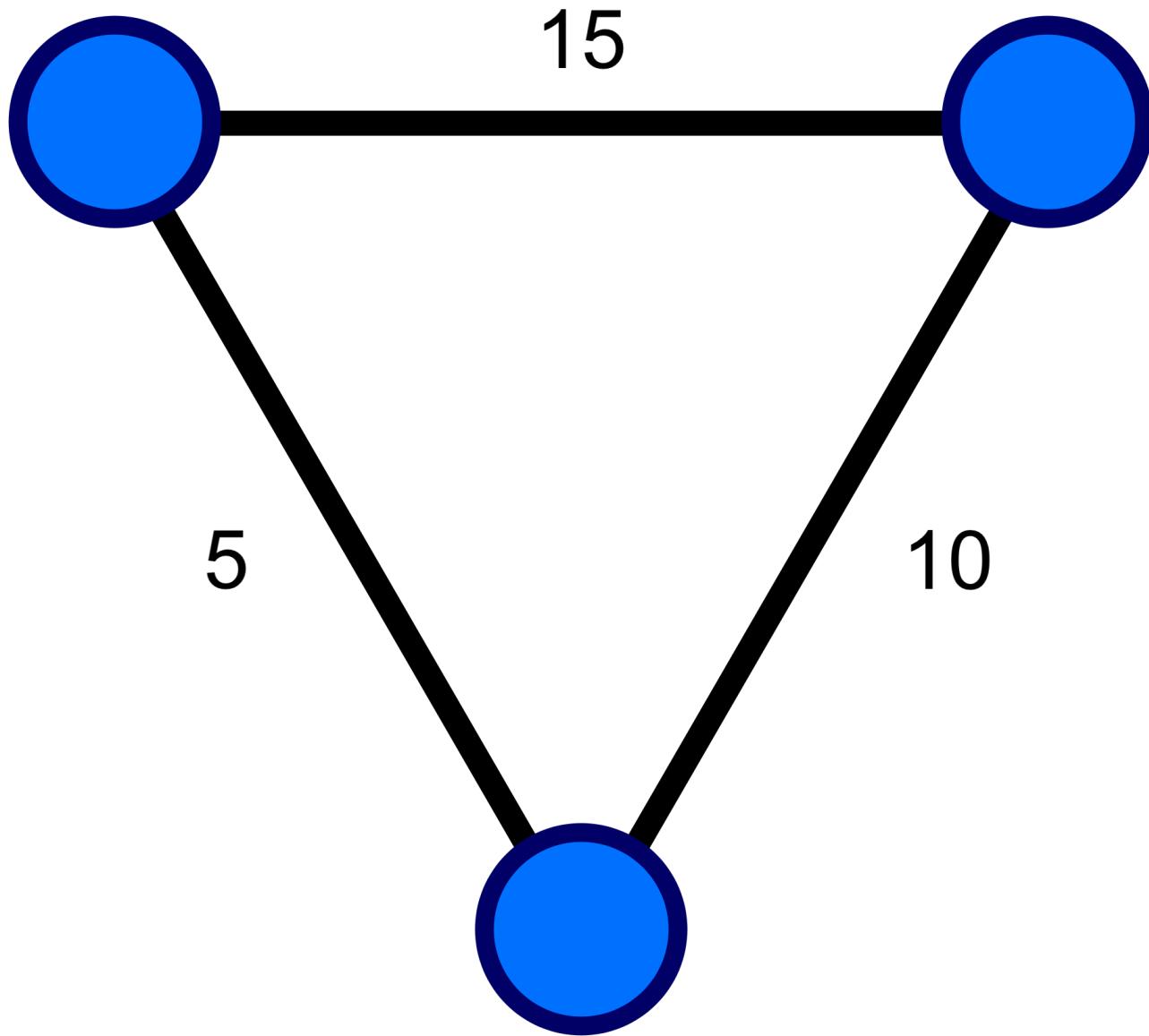
Non-orienté (*undirected graph*)



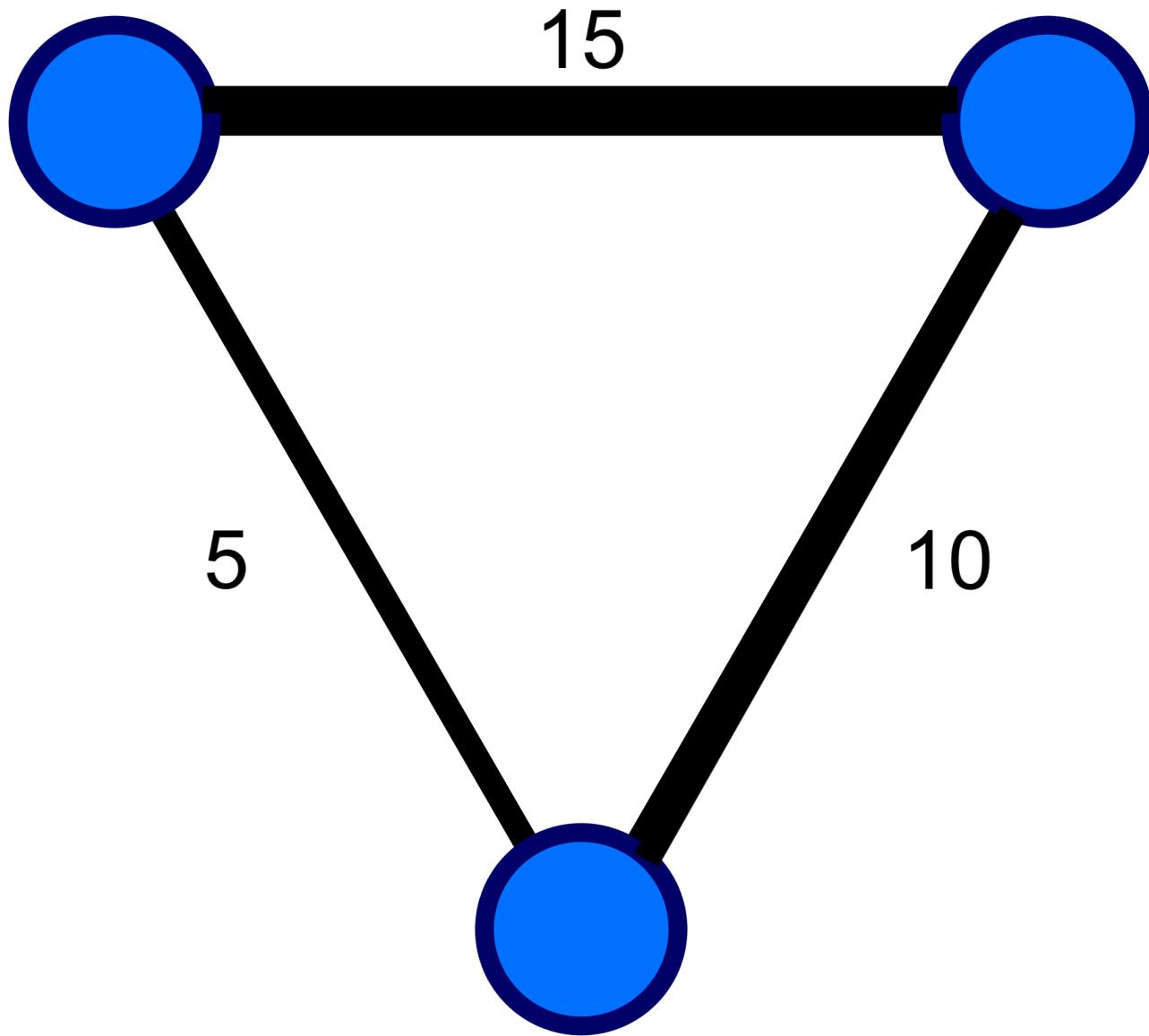
Orienté (*directed graph*)



Valué (*weighted* ou *valued* graph)



Valué (*weighted* ou *valued* graph)

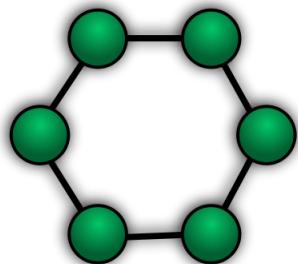


Les Cliques

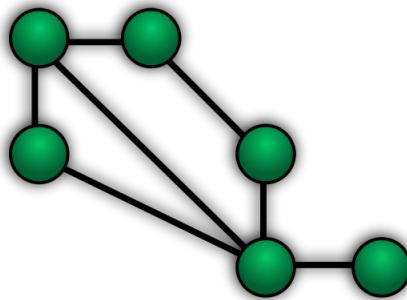
Les réseaux forment des *cliques* (*clusters*) ou *communautés*. Ces clusters peuvent s'organiser selon différents types.

Les types de réseaux

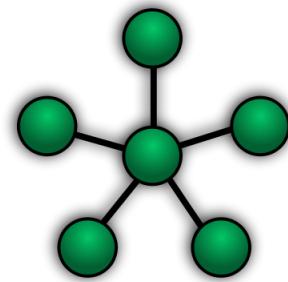
- Cercle (*ring*)
- Filet (de poisson) (*mesh*)
- Étoile (*star*)
- Complet (*fully connected*)
- Ligne (*line*)
- Arbre (*tree*)
- Vide (sans lien)
- Bus (*bus*)



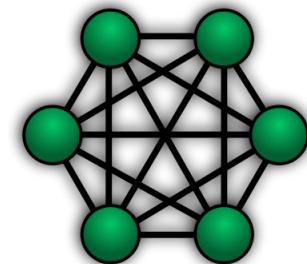
Ring



Mesh



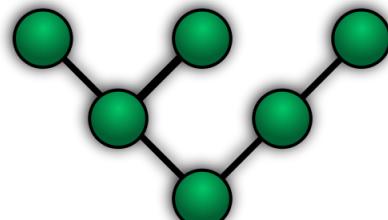
Star



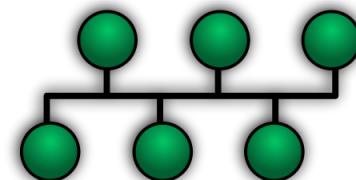
Fully Connected



Line



Tree



Bus

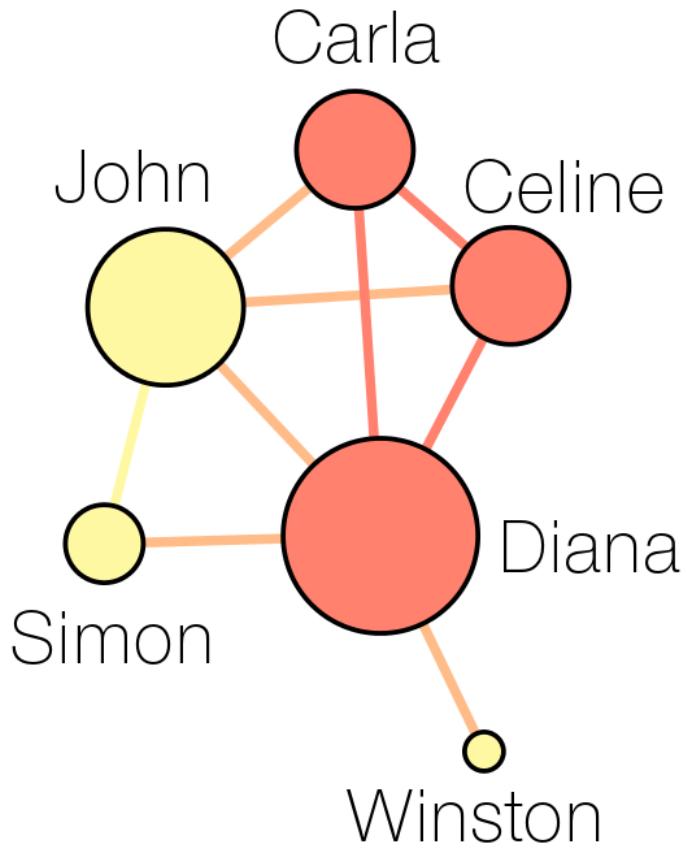
Construction du réseau

Construction

Un graphe peut se présenter sous trois formes

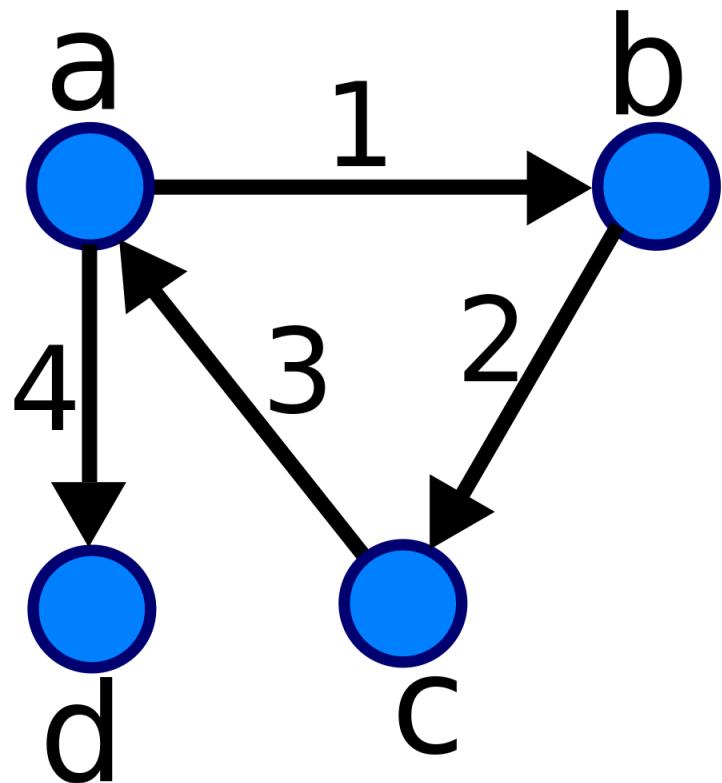
- Le graphe lui-même
- Une matrice
- Une liste de nœuds et d'arêtes

Deux fichiers



Nodes	Edges
Id,Label,Gender	Source,Target
1,John,1	1,2
2,Carla,2	1,3
3,Simon,1	1,4
4,Celine,2	1,6
5,Winston,1	2,4
6,Diana,2	2,6
	3,6
	4,6
	5,6

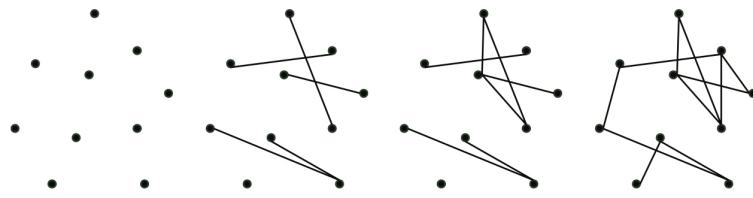
Une matrice



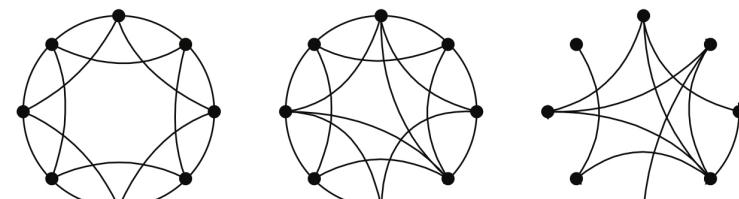
	1	2	3	4
a	1	0	-1	1
b	-1	1	0	0
c	0	-1	1	0
d	0	0	0	-1

Les modèles de réseaux

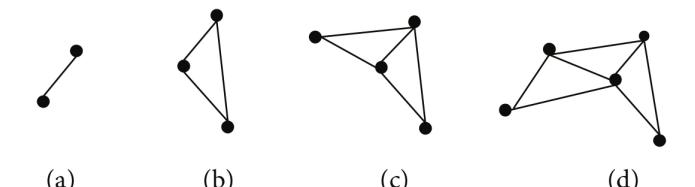
- *Random Network* (RN)
- *Small-World Network* (SWM)
- *Scale-Free Network* (SFN), ou "sans échelle"



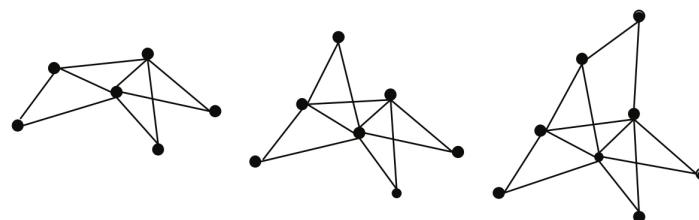
(a) (b) (c) (d)
(1) ER random network



(a) (b) (c)
(2) WS small world network



(a) (b) (c) (d)

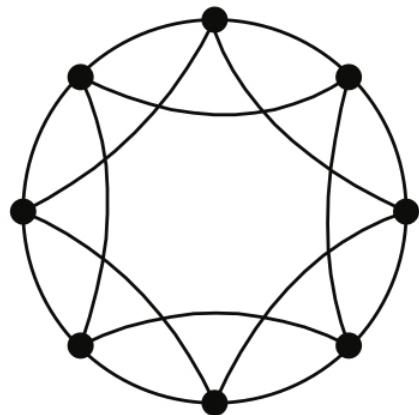


(e) (f) (g)
(3) BA scale-free network

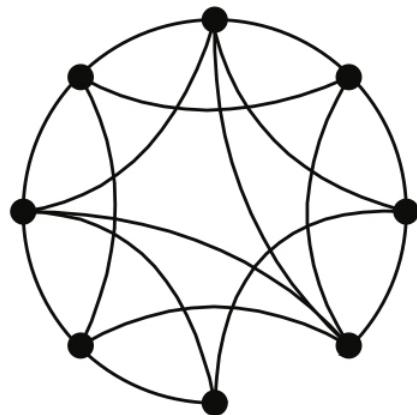
Source Xu, Li, Liu 2014

Small-World Network ou petit monde (SWM)

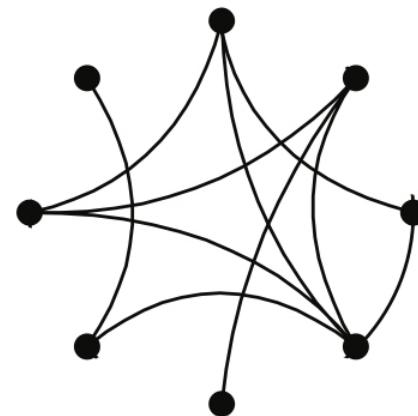
- Distribution gaussienne
- *Top-down*: je pars de toutes les combinaison possibles et j'en retire petit à petit



(a)



(b)



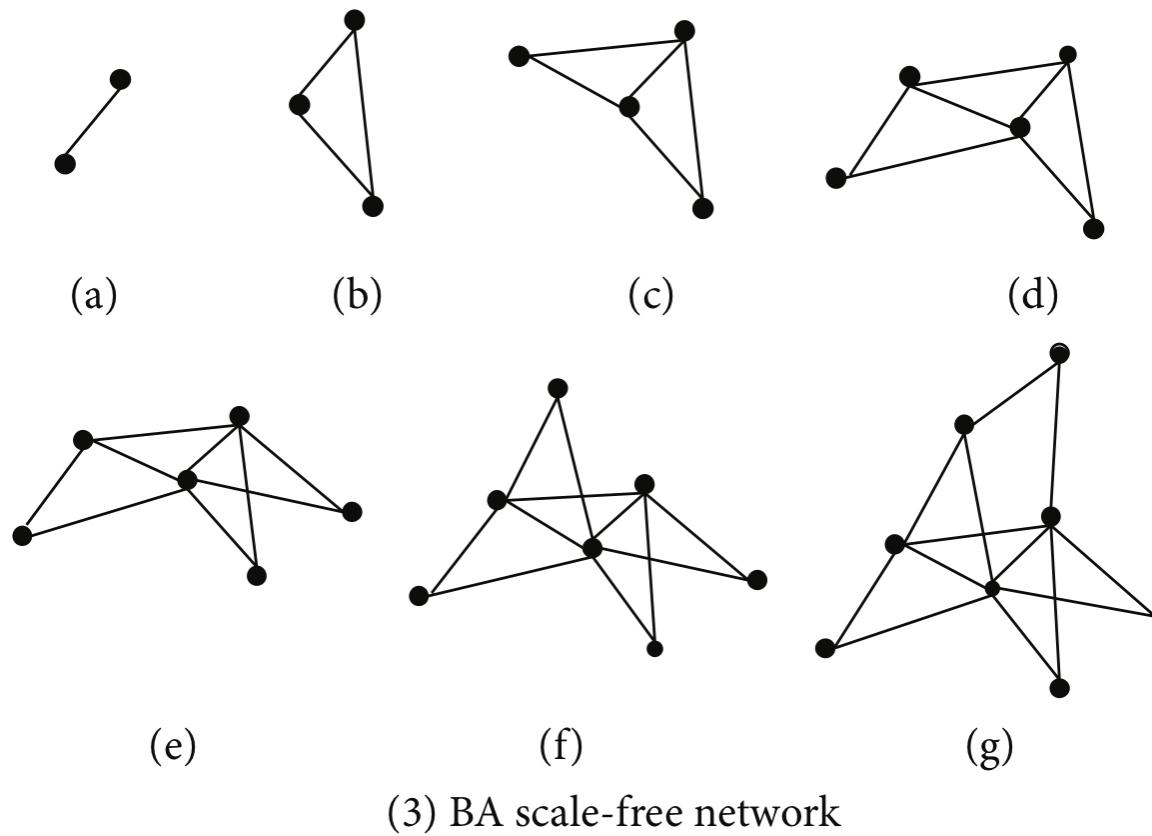
(c)

(2) WS small world network

Source Xu, Li, Liu 2014

Scale-Free Network ou sans échelle (SFN)

- Loi de puissance (peu de nœuds concentre un grand nombre de lien)
- *Bottom-up*: je pars de rien, et je rajoute petit à petit les arêtes



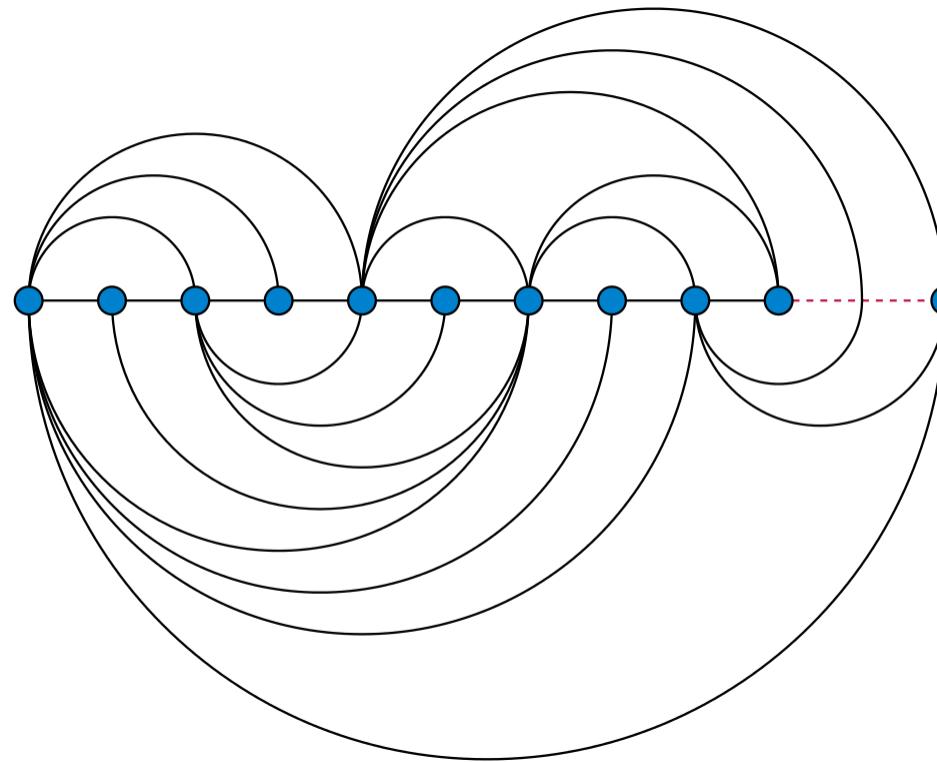
Tracé de graphe

Il va falloir représenter des graphes dans le plan. Pour cela il existe plusieurs méthodes:

- L'*arc diagram*
- Le *circular layout*
- Le *Force-based layout* (le plus connu)
- Bien d'autres...

Arc diagram

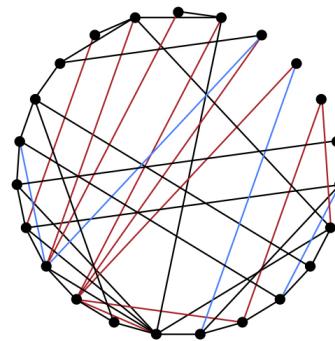
- Date des années 1960
- Les nœuds sont placées sur une seule et même ligne, avec les arêtes sous la forme de demi-cercles sur/sous la ligne
- Possible avec un nombre limité de données



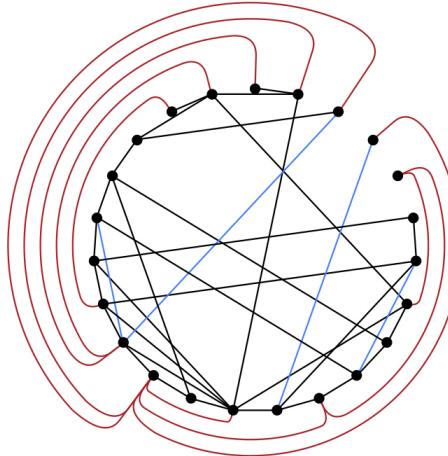
Source: [Wikipedia](#)

Circular layout

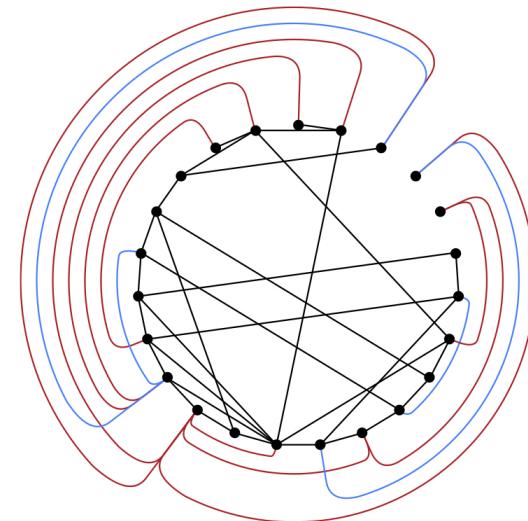
- Représentation des nœuds sur un cercle avec un espace régulier
- Il s'agit de minimiser le nombre de croisement des arêtes.
- Pour optimiser ces croisements, on tente de développer des alternatives, comme le *two-Sided Circular Graph* (cf. *infra*)



(a) One-sided layout with 125 crossings.



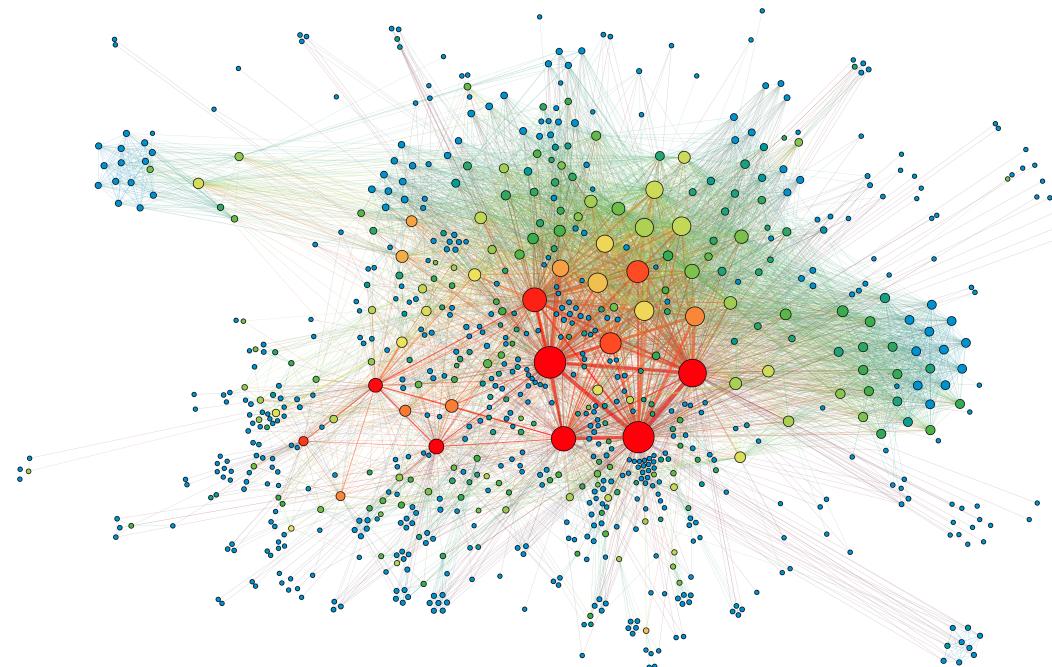
(b) Two-sided layout for $k = 0$ with 48 crossings.



(c) Two-sided layout for $k = 1$ with 30 crossings.

Force-based layout

L'objectif est de positionner les nœuds d'un graphe pour faciliter sa visualisation en utilisant un système de force appliqué entre les nœuds et les arcs/arêtes.



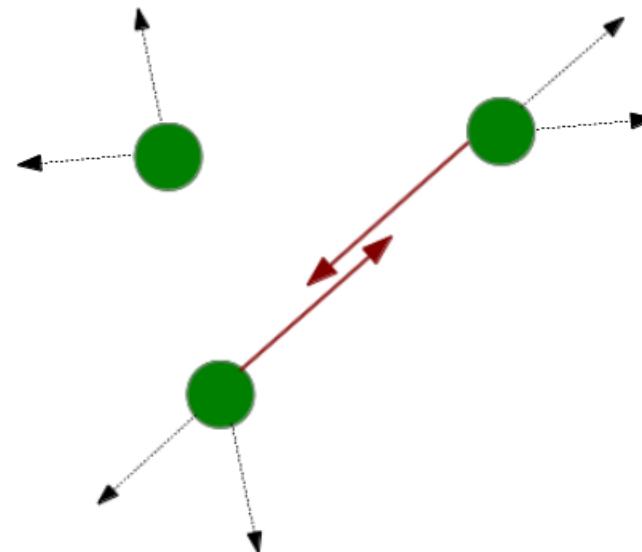
Source: [wikipedia](#)

Pour cela, on va utiliser des algorythmes qui vont:

1. Repousser les nœuds entre eux, comme des aimants, les plus éloignés se repoussant le moins



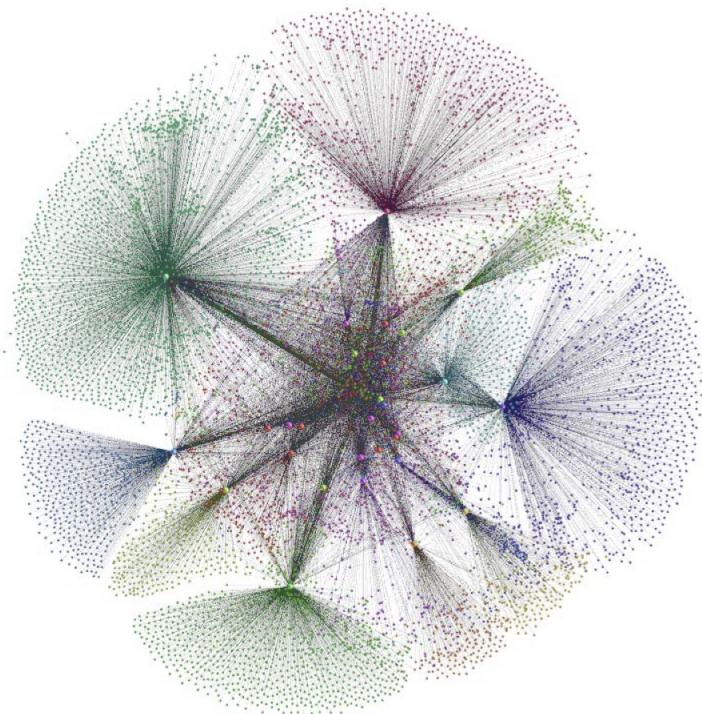
2. Les liens fonctionnent comme des ressorts



Source Rémi Damлencour

Fruchterman and Reingold Algorithm

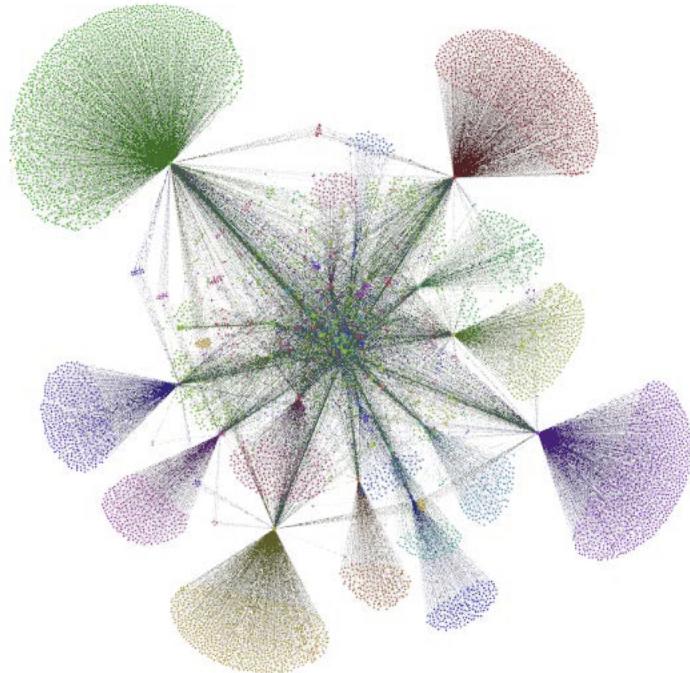
- *Fruchterman and Reingold Algorithm*, le plus ancien, presque obsolète



Lily Popova Zhuhadar, "Cyberlearners and learning resources", ACM
Transactions on the Web, 2012 [lien](#)

Yifan Hu Algorithm

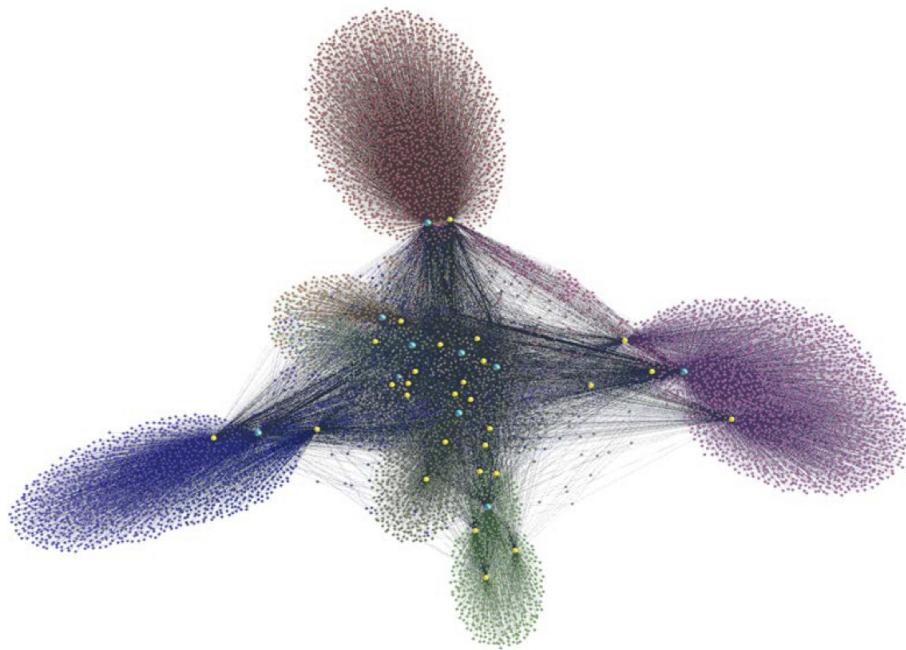
Pour une modélisation rapide de petits réseaux (max 100 000 de nœuds)



Lily Popova Zuhadar, "Cyberlearners and learning resources", ACM
Transactions on the Web, 2012 [lien](#)

Force Atlas 2 Algorithm

Pour analyse de complémentarités et les grands réseaux (1 000 000 de nœuds)

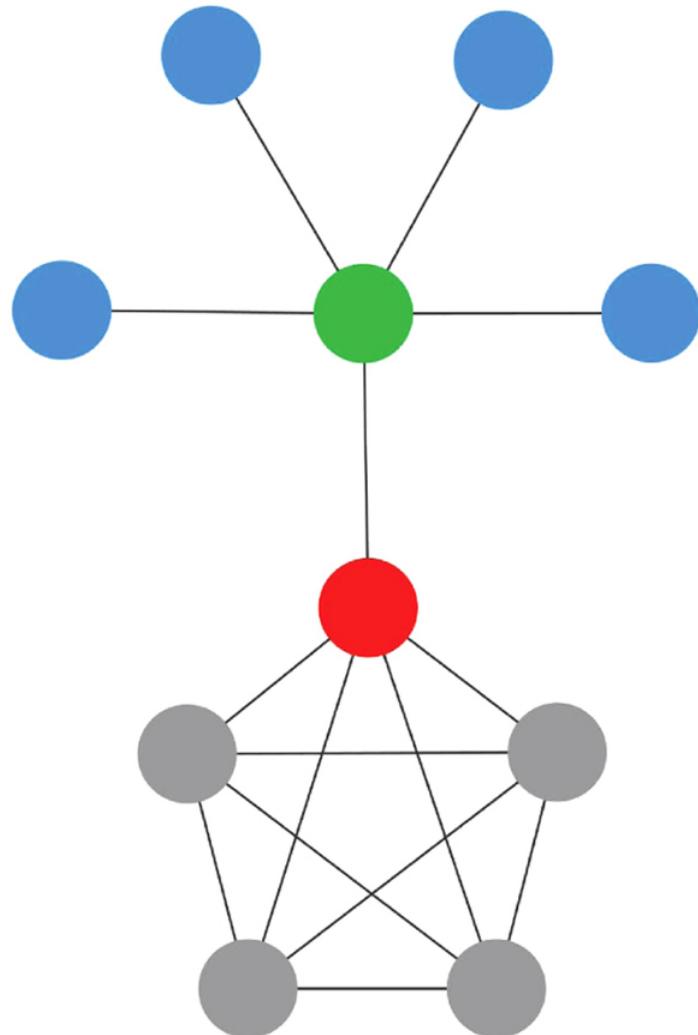


Lily Popova Zhuhadar, "Cyberlearners and learning resources", ACM
Transactions on the Web, 2012 [lien](#)

Analyse

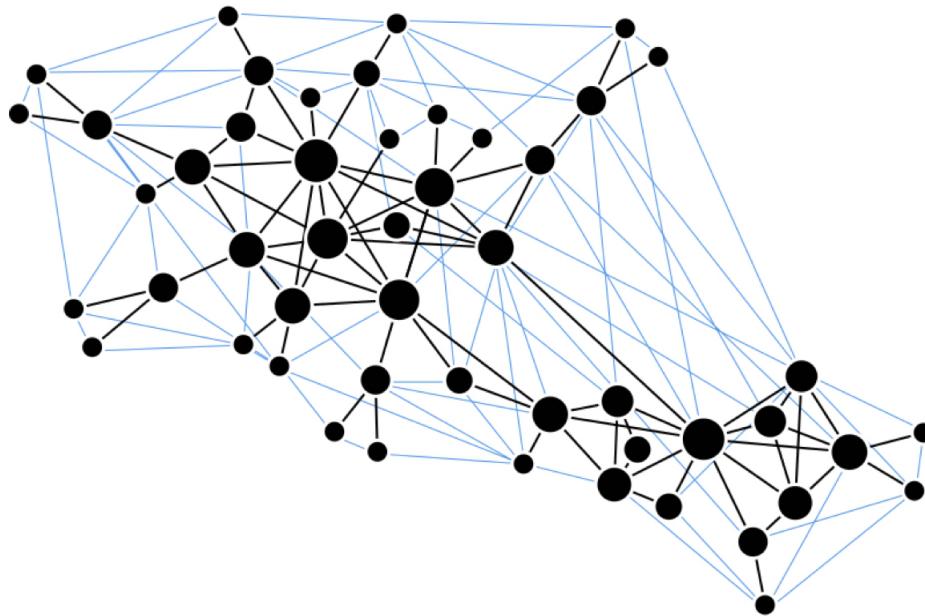
Robustesse

Robustesse (*robustness*): capaciter du graphe à rester identique malgré des perturbations (enlever un nœud)



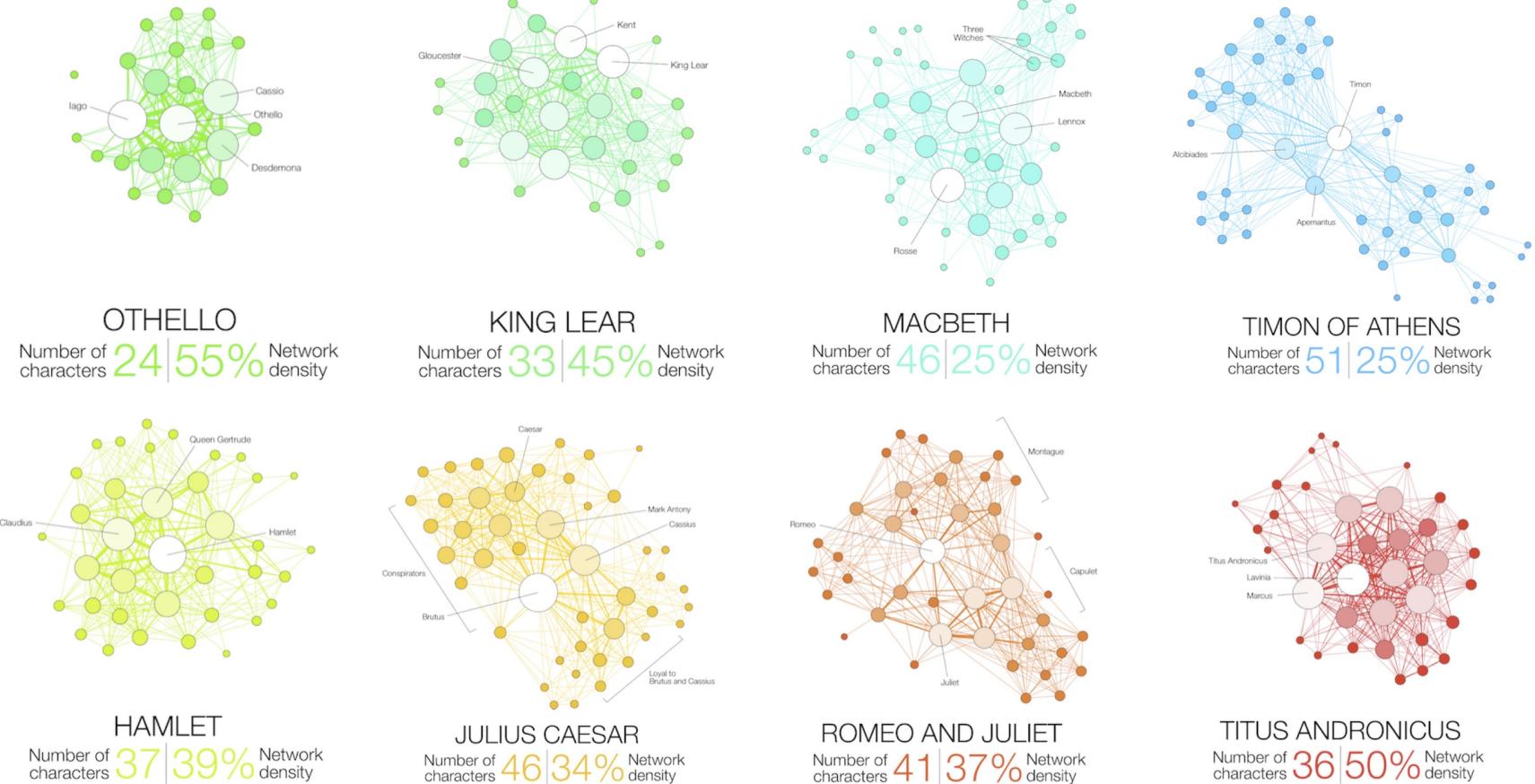
Densité

Densité (*density*): la proportion de liens dans un réseau relativement au total des liens possibles.



M. Grandjean et M. Jacomy. "Translating Networks: Assessing Correspondence Between Network Visualisation and Analytics". *Digital Humanities*, 2019, Utrecht. [lien](#)

Shakespeare: densités



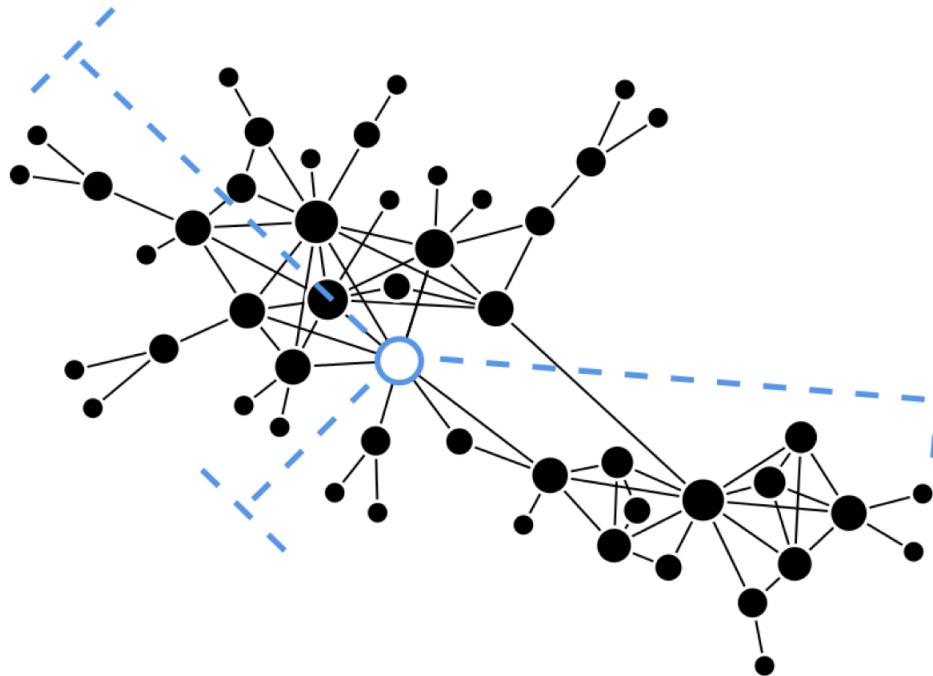
Source: [Grandjean 2015](#)

Centralité

Centralité (*centrality*) : capturer la notion d'importance dans un graphe, en identifiant les sommets les plus significatifs

Proximité

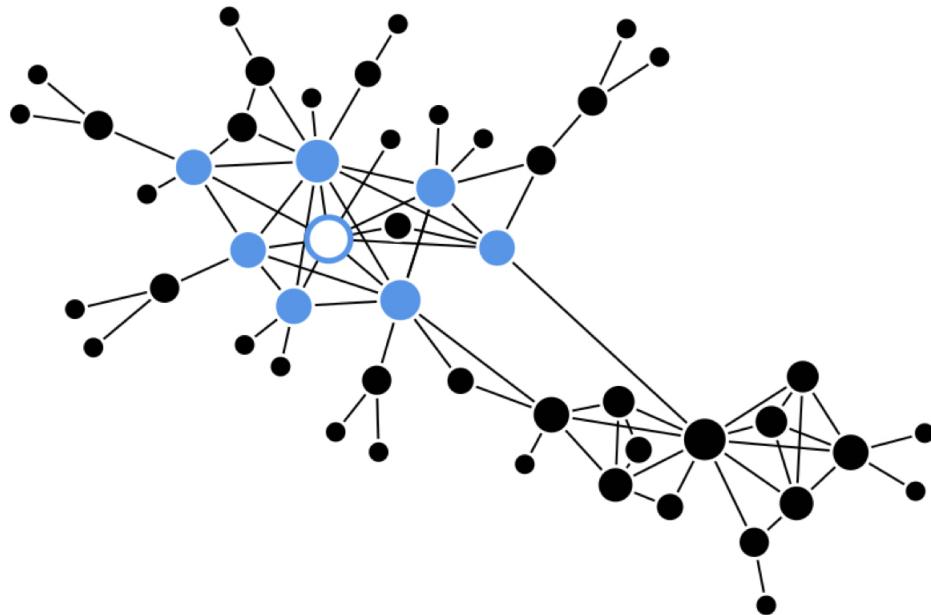
Centralité de proximité: Distance moyenne du nœud à tous les autres nœuds (*Closeness*)



M. Grandjean et M. Jacomy. "Translating Networks: Assessing Correspondence Between Network Visualisation and Analytics". *Digital Humanities*, 2019, Utrecht. [lien](#)

Prestige

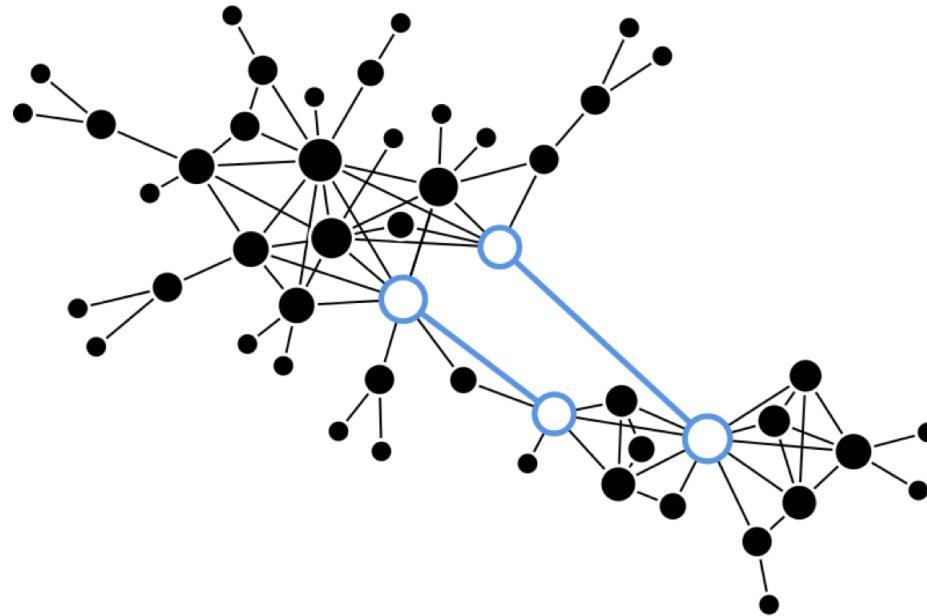
Centralité de vecteurs propres: Score d'autorité attribué à un nœud en fonction du score de ses voisins. (*Eigenvector*) C'est l'algorythme de google.



M. Grandjean et M. Jacomy. "Translating Networks: Assessing Correspondence Between Network Visualisation and Analytics". *Digital Humanities*, 2019, Utrecht. [lien](#)

Intermédiairité

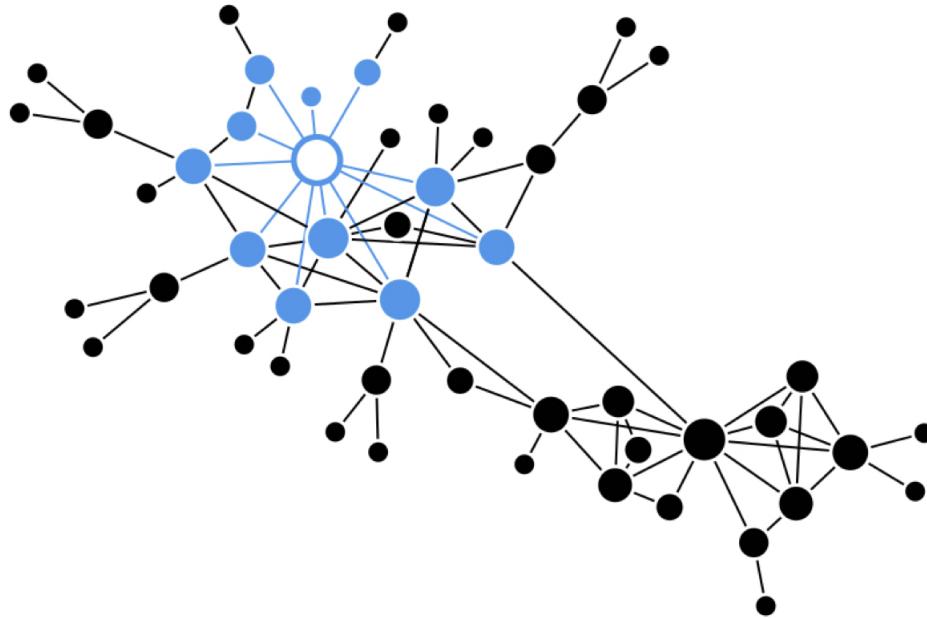
Centralité d'intermédiairité: Nombre de fois que le nœud se trouve sur le plus court chemin entre deux autres nœuds (*Betweenness*)



M. Grandjean et M. Jacomy. "Translating Networks: Assessing Correspondence Between Network Visualisation and Analytics". *Digital Humanities*, 2019, Utrecht. [lien](#)

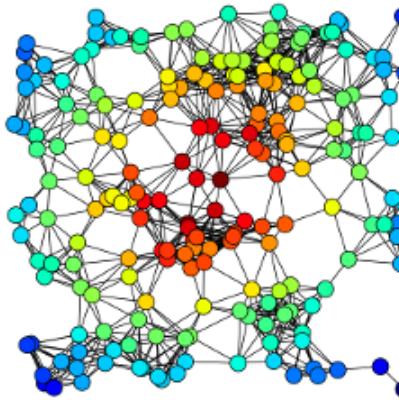
Degré

Centralité de degré: Nombre de connexions du nœud (*Degree*)

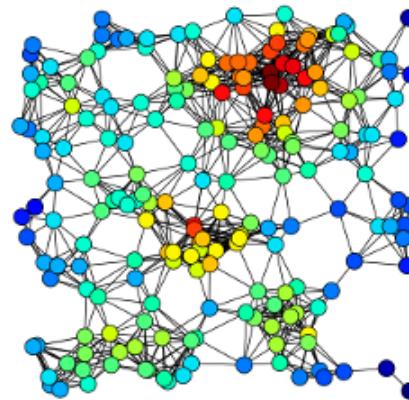


M. Grandjean et M. Jacomy. "Translating Networks: Assessing Correspondence Between Network Visualisation and Analytics". *Digital Humanities*, 2019, Utrecht. [lien](#)

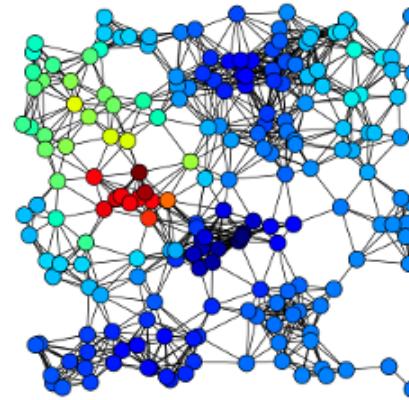
A) Betweenness centrality, B) Closeness centrality, C) Eigenvector centrality, D) Degree centrality, E) Harmonic centrality et F) Katz centrality



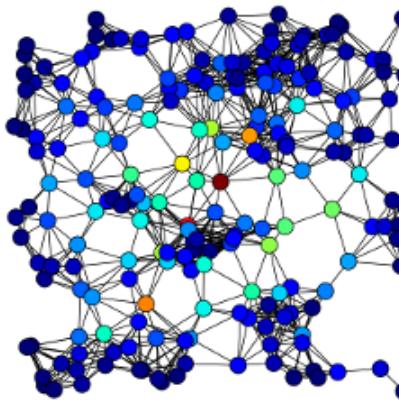
B



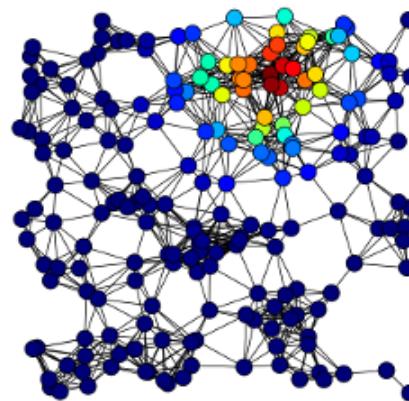
D



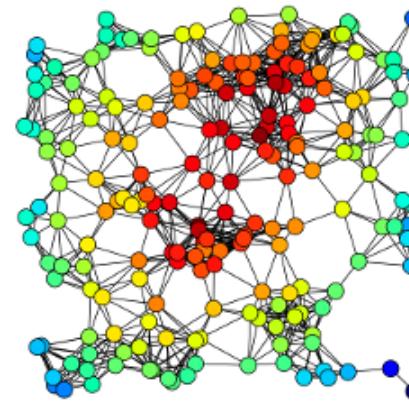
F



A



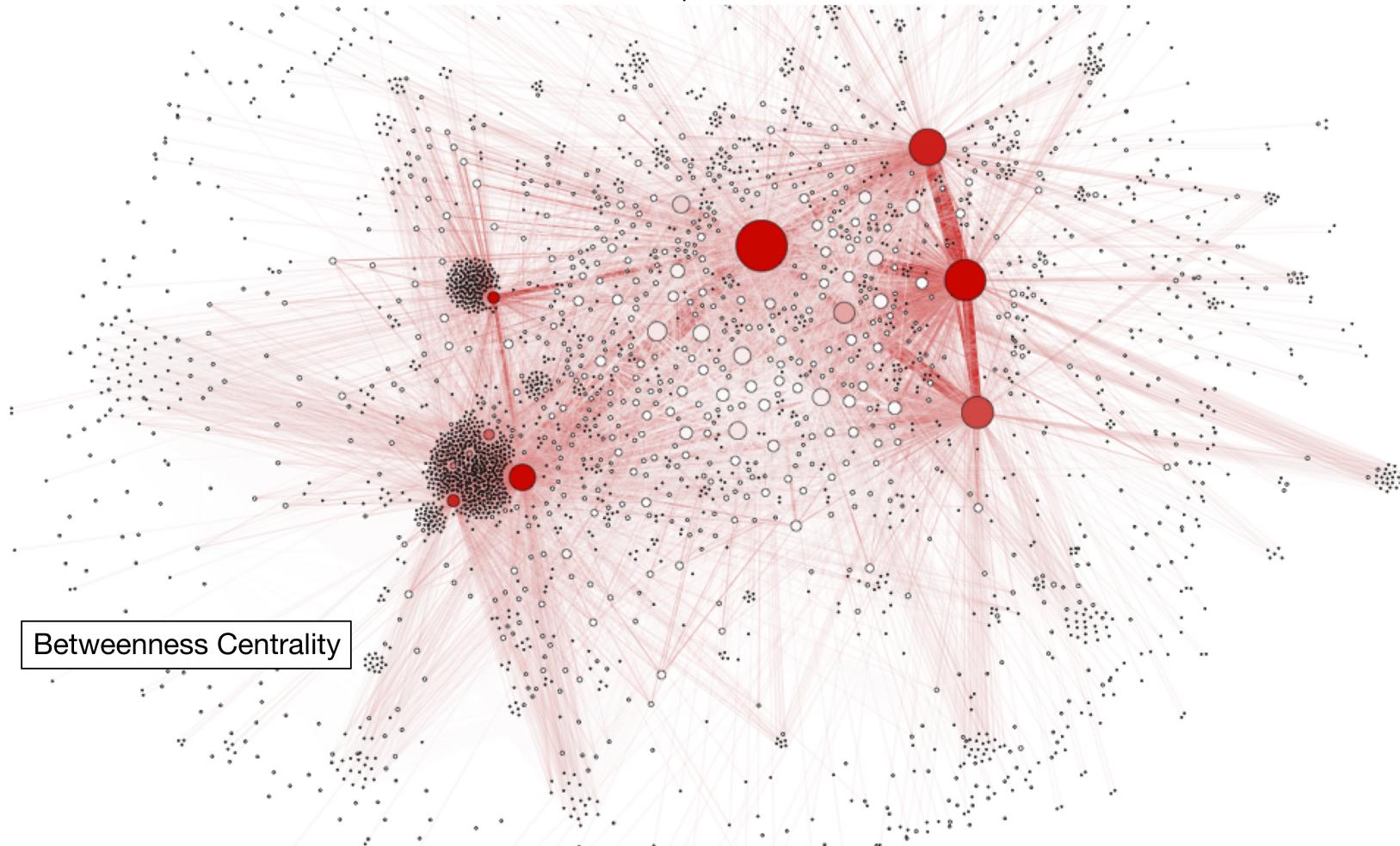
C



E

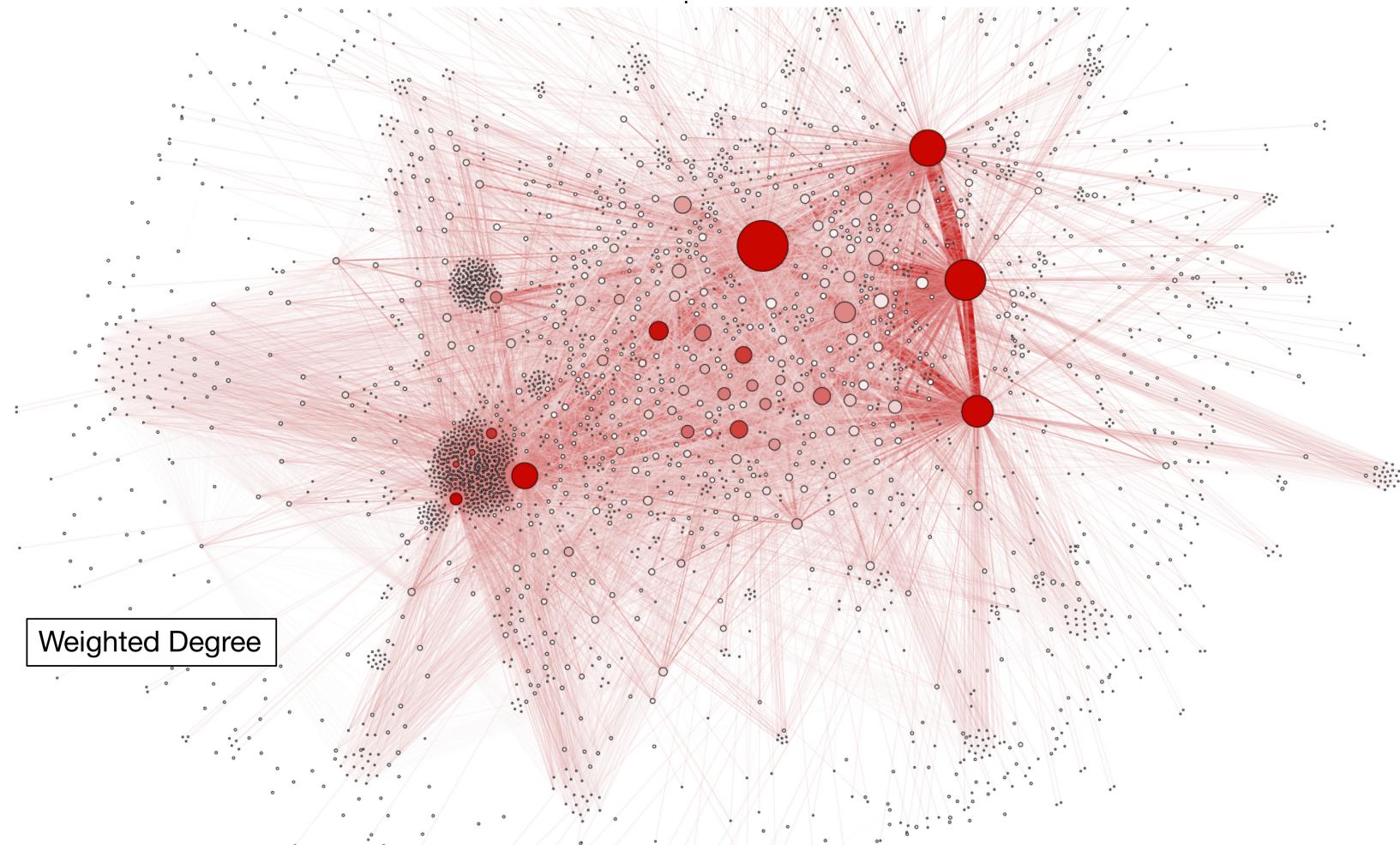
Source: [Wikipedia](#)

Ligue des nations: *betweenness centrality*



Source: [Grandjean 2015](#)

Ligue des nations: *weighted degree centrality*



Source: [Grandjean 2015](#)