

Introduction à l'analyse des réseaux sociaux

MSH Lyon / Saint-Etienne

Maxime Cornet - Institut Polytechnique de Paris

27 Octobre 2021

Introduction:

- I Qu'est ce qu'un réseau? Terminologie & Théorie des graphes
- II Pourquoi étudier la structure des interactions? Origine du concept
- III Calculs & métriques de base
- IV Exemples d'application

I) Qu'est-ce qu'un réseau social?

« l'analyse des réseaux sociaux est un ensemble de sous méthodes pour analyser des ensembles de relations sociales, considérées elles-mêmes comme émergeant des interactions entre des personnes ou des groupes »

- Degenne, Forsé, 1994

I) Qu'est-ce qu'un réseau social?

- Un outil d'analyse
- Prise en compte de l'intégralité des relations interactionnelles données, pour observer la structure relationnelle qui émerge de ces relations
- Dans une temporalité propre
- Survenant dans un contexte délimité

I) Comment ça se formalise?

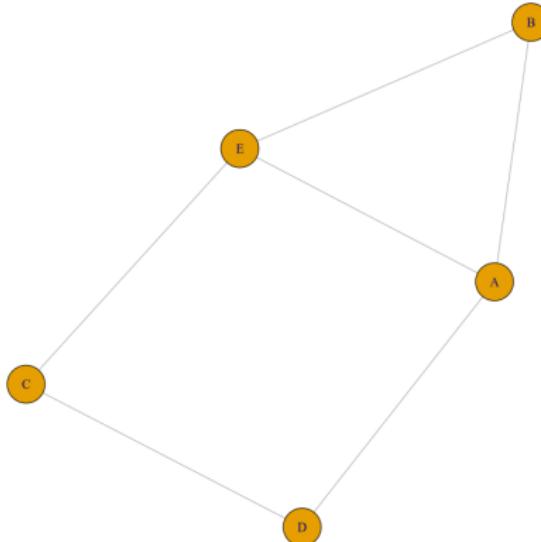


Figure: Exemple de réseau

I) Comment ça se formalise?

	A	B	C	D	E	
A	0	1	0	1	1	
B	1	0	0	0	0	1
C	0	0	0	1	1	1
D	1	0	1	0	0	0
E	1	1	1	0	0	0

Figure: Matrice du réseau vu en slide précédente

I) L'analyse de réseaux

- Théorie des graphes: une branche des mathématiques remontant au « problème des sept ponts de Konigsberg » d'Euler.
- Noeuds (nodes): vertex; acteurs
- Arrête (edge): arcs; liens; relations
- Plusieurs types: réseaux complets; réseaux personnels réseaux techniques
- Plusieurs formes: dirigés / non dirigés; one-mode / two mode

I) Réseaux dirigés

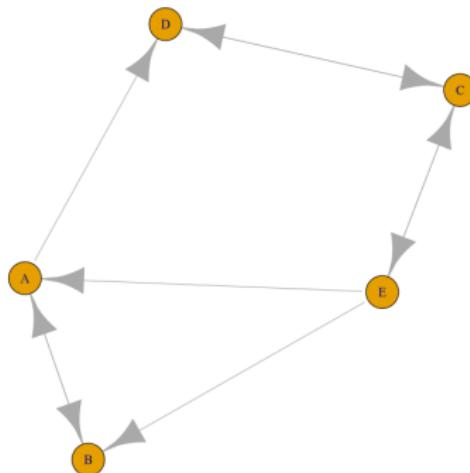


Figure: Exemple de réseau dirigé

I) Réseaux dirigés

	A	B	C	D	E
A	0	1	0	1	0
B	1	0	0	0	0
C	0	0	0	1	1
D	0	0	1	0	0
E	1	1	1	0	0

Figure: Matrice du réseau vu en slide précédente

I) Collecter des données de réseau

- Variabilité des sources (webscrapping, questionnaire, entretiens, archives)
- Variabilité des modes de collecte (rappel libre, liste à coche, sociogramme, etc.)
- Biais attendus et spécifiques aux dispositif de collecte
- Attention aux données manquantes: problèmes spécifique pour les données de réseau.

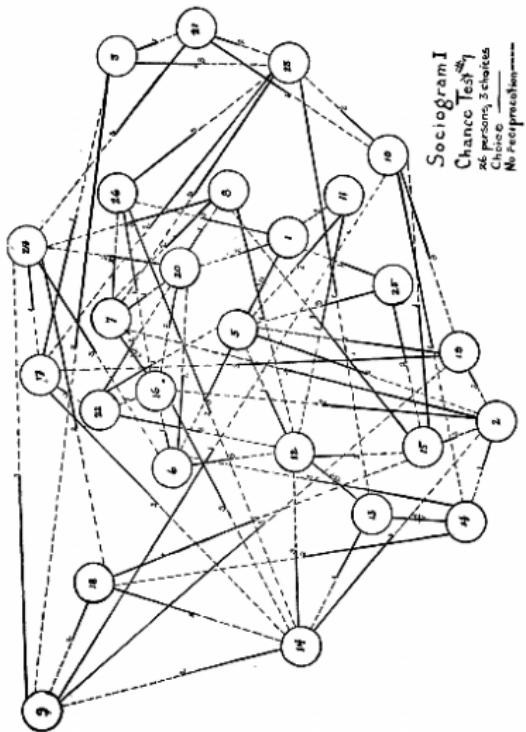
I) Délimiter un réseau

- Le réseau résulte d'un ensemble de décisions méthodologiques qui doivent être prises consciemment en rapport avec les hypothèses de départ.
- Qui sont les noeuds? Quelles interactions relever?
- Ou commence et s'arrête le réseau? Spatialement et temporellement?
- L'analyse de la structure n'a de sens qu'en gardant ces aspects méthodologiques et réflexifs en tête.

II) Une structure qui émerge des interactions?

- Georg Simmel (1905): la triade est l'unité d'analyse fondamentale en sociologie.
- La dynamique sociale de trois acteurs dans une triade est qualitativement différente de ce qui est issu de l'observation d'une dyade ou d'un individu isolé.
- Les comportements et phénomènes sociaux ne peuvent pas être réduits à des dyades ou individus.

II) Méthodes anciennes



Première représentation que l'on connaît (celle en graphe), on la doit à Moreno: Ille sociogramme.
- Moreno, 1938

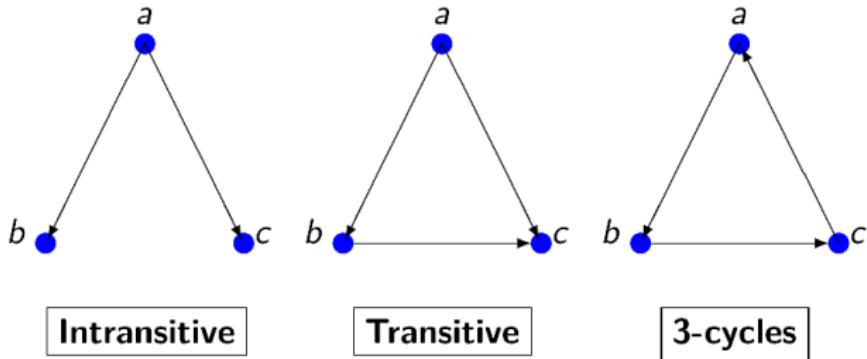
II) Dynamiques sociales

- J. A. Barnes (1905): S'appuie en partie sur les travaux de Moreno, et introduit la notion de "réseau", et la notion de multiplicité de réseau entremêlés.
- "I find it convenient to talk of a social field of this kind as a network (...) We can of course think of the whole of social life as generating a network of this kind".
- "So far we have looked at Bremnes society as an isolated object of study. (...) In reality, Bremnes is not an isolated society, and there is the large descriptive and analytical problem of understanding the relationship between bremnes and neighbouring parishes, and between it and the Norwegian state."

II) Etudier les interactions

Abbott (1997, 2001): Les théories sociologiques s'intéressent aux acteurs et à leurs interactions, alors que les méthodes sociologiques s'intéressent aux variables et à leurs interactions .Il y a là une incohérence entre théorie et méthode, qui ne peuvent pas se nourrir l'une l'autre.

II) Transitivité:



II) Nature dynamique des interactions:

- La plupart des réseaux sociaux observés empiriquement ont un degré élevé de transitivité. (Beaucoup plus élevé que dans les réseaux aléatoires)
- La transitivité est une caractéristique essentielle des relations humaines : les êtres humains se lient entre eux par des intermédiaires.
- Granovetter (1973): La triade intransitive est instable, on peut prédire qu'un lien se formera par transitivité tôt ou tard.

III) Métriques structurelles et définitions: globales

- Densité: Mesurer la cohésion (globale) : Le rapport entre le nombre de liens qui existent et ceux qui pourraient exister.
- Diamètre: la distance entre deux nœuds les plus distants dans un réseau.
- Distance moyenne: distance moyenne entre toutes les paires de nœuds.
- Rayon: la plus petite distance à laquelle puisse se trouver un nœud de tous les autres.

III) Métriques structurelles et définitions: globales

- Transitivité: L'indice de transitivité globale d'un réseau: nombre triades transitives / nombre triades.
- Il est égal à 1 si tous les nœuds sont liés à tous les autres nœuds (connectivité complète).
- Dans des réseaux aléatoires, il excède rarement 0.2. Dans les réseaux empiriques, il est souvent compris entre 0.3 et 0.6
- Variante locale: Le coefficient de clustering (ou coefficient d'agglomération), est une mesure de la cohésion dans le voisinage d'un nœud (combien de mes amis sont amis entre eux).

III) Métriques structurelles et définitions:

- Réciprocité: dans les réseaux dirigés, taux de liens qui vont dans les deux sens.
- Nombre de cliques: Une clique est un sous-ensemble de nœuds où toutes les paires de nœuds existants sont connectées. (ex: d'implication structurelles: Co-appartenance comme indicateur de cohésion; Alba & Kadushin, 1976). Taille moyenne des cliques et plus grande clique.
- Excentricité: la distance depuis un nœud de départ vers le nœud le plus loin dans le réseau.

III) Centralité:

- On tente de répondre à la question, quel rôle occupe quel noeud dans le réseau? -> Calcul de propriétés structurelles des noeuds.
- Les acteurs ont des caractéristiques propres, qui dépendent de la structure globale du réseau et de leur positon dans celle ci.
- Plusieurs types de centralité, à mobiliser en fonction des hypothèses de départ.

III) Centralité de degrés:

- Centralité de degré: Qui sont les nœuds les plus « actifs »?
On compte tout simplement le nombre de lien qui relient un noeud à d'autres.
- Comparaisons de centralité possible seulement si les réseaux sont de même taille : degré = 8 veut dire être connecté à 80% s'il y a 10 acteurs, à 20% s'il n'y en a que 40. (utiliser le calcul normalisé)
- Degré entrant / sortants: pour les graphes connectés: nombre de liens émis / reçus.

III) Centralité:

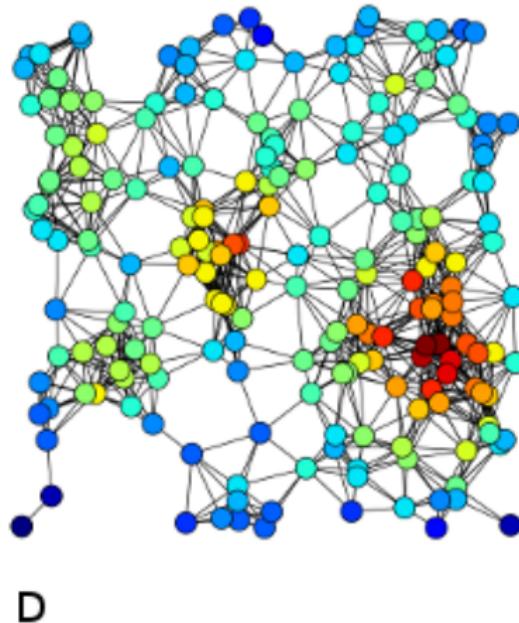
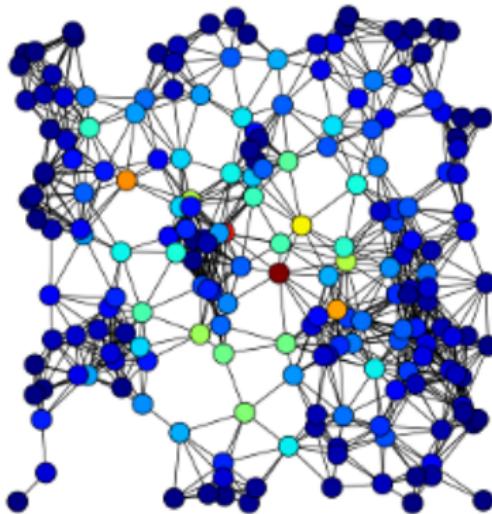


Figure: Centralité de degrés. Rouge = plus élevée. source: wikipédia

III) Centralité d'intermédiarité:

- La centralité d'intermédiarité identifie les gatekeepers dans le réseau. Davantage de chemins passent par ces acteurs, qui peuvent passer l'information à d'autres dans le réseau (ou la stopper).
- La centralité d'intermédiarité est parfois une mesure plus pertinente de l'« importance » dans un réseau. Les acteurs avec plus grande centralité d'intermédiarité:
 - Connectent des groupes différents dans un réseau, et constituent parfois le seul lien entre eux
 - Contrôlent les flux de communication dans le réseau
 - Peuvent agir en intermédiaires
- Calcul: Il faut prendre en compte non seulement les liens directs de l'acteur i, mais tout le réseau. On compte le nombre de plus courts chemins entre toute paire d'acteurs k et j, et on prend ceux qui passent par i.

III) Centralité:



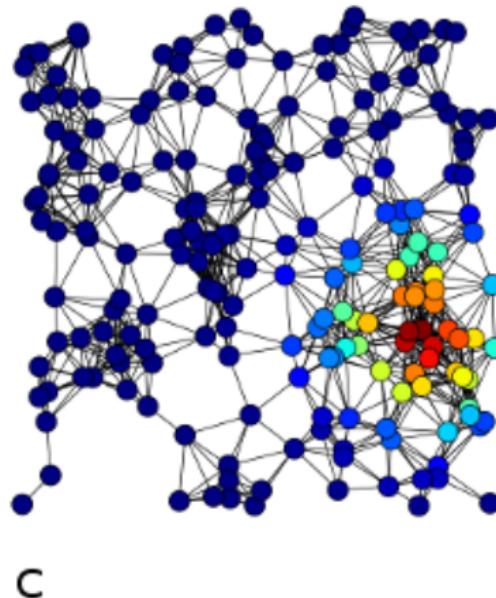
A

Figure: Centralité d'intermédiarité. Rouge = plus élevée. source: wikipédia

III) Centralité de vecteur propre (eigenvalue):

- Vecteur propre: Donne plus de poids aux acteurs avec le plus grand nombre de liens. Mais elle donne un poids plus élevé aux contacts qui sont eux-mêmes centraux dans le réseau.
- Elle effectue une pondération des contacts sur la base de leur centralité. Ce faisant, elle tient compte de la structure du réseau dans son ensemble.
- La centralité eigenvector peut être vue comme une version récursive de la centralité de degré :
 - ① Commencez en assignant centralité 1 à tous les nœuds
 - ② Recalcullez le résultat pour chaque nœud comme somme pondérée des centralités de tous les nœuds dans son voisinage.
 - ③ Normalisez en divisant chaque valeur par le maximum
 - ④ Répétez les étapes 2 et 3 jusqu'à ce que les valeurs arrêtent de changer.

III) Centralité:



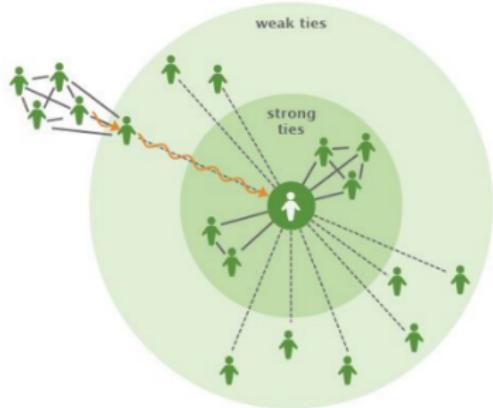
C

Figure: Centralité de vecteur propre. Rouge = plus élevée. source: wikipédia

III) Centralisation:

- Dans quelle mesure le réseau est dominé par un nœud central (ou peu de nœuds centraux)?
- On compare la centralité du nœud le plus central à la centralité des autres nœuds. On normalise en divisant par la centralité maximale d'un réseau de la même taille.
- Le résultat varie entre 0 et 1 (réseau en étoile). Elle tends à varier dans le temps.

IV) Quelques exemples d'usage des réseaux:



- Granovetter (1973, 1974): importance des relations pour trouver un emploi
- Liens forts moins efficaces que liens faibles
- Un défi pour la théorie sociologique de l'époque, qui insistait sur les bienfaits de la cohésion

IV) Quelques exemples d'usage des réseaux:

- Conceptuellement, Granovetter étudie les chaînes d'accès à une ressources mobilisées par un acteur, en les contextualisant grâce à son réseau personnel
- Notion liens forts / liens faibles: attributs des liens, ici liés à la fréquence de l'interaction.
- Nombreuses décisions en lien avec les théories mobilisées par Granovetter: réseaux personnels, chaîne interactionnelles, etc.

IV) Quelques exemples d'usage des réseaux:

- Martina Morris (1995): étudie la diffusion du VIH dans les années 90, en particulier les effets de différentes "*culturally specific sexual network structure*" sur la diffusion du virus.
- Elle trouve que la forte présence de relations co-occurentes dans un réseau réduit la probabilité de survenue d'une vague épidémique, mais augmente fortement la diffusion en cas de survenue, par rapport à la monogamie séquentielle.
- Elle trouve aussi que la configuration "dissortative", où les noeuds les plus actifs ont des relations avec les noeuds les moins actifs est la plus favorable à la diffusion du virus.

IV) Quelques exemples d'usage des réseaux:

M. Morris, M. Kretzschmar / Social Networks 17 (1995) 299–318

301

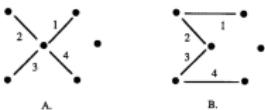


Fig. 1. Two different contact graphs with equal numbers of nonisolate nodes and links, and therefore equal values for the concurrency measure κ_1 (defined in Eq. (1)).

- Importance de croiser les mesures
- Le taux de co-occurrence des relations est le même dans les deux réseaux, mais les schémas de diffusion sont très différents

IV) Quelques exemples d'usage des réseaux:

- Morris travaille sur des modèles relationnels de diffusion: ce ne sont pas des réseaux empiriques, mais des réseaux simulés.
- Les paramètres des modèles sont tirés d'études réelles, mais la structure relationnelle est entièrement artificielle.
- Morris utilise ces différents modèles de structure pour tester des hypothèses de diffusion. Elle définit aussi les sous-structures qui lui semblent importantes pour tester ses hypothèses.

IV) Quelques exemples d'usage des réseaux:

- Gaumont, Panahi, Chavalarias (2018): Étudient les réseaux socio-sémantiques politiques du twitter francophone pendant l'élection présidentielle française de 2017.
- Se focalisent sur les retweet de comptes politiques, et en infèrent une appartenance politique.
- Ils étudient la manière dont la structure du réseau de diffusion sur twitter évoluent durant toute la campagne. Ils traitent donc de l'aspect évolutif et longitudinal des réseaux.

IV) Quelques exemples d'usage des réseaux:



Figure: Avant le début de la campagne

IV) Quelques exemples d'usage des réseaux:

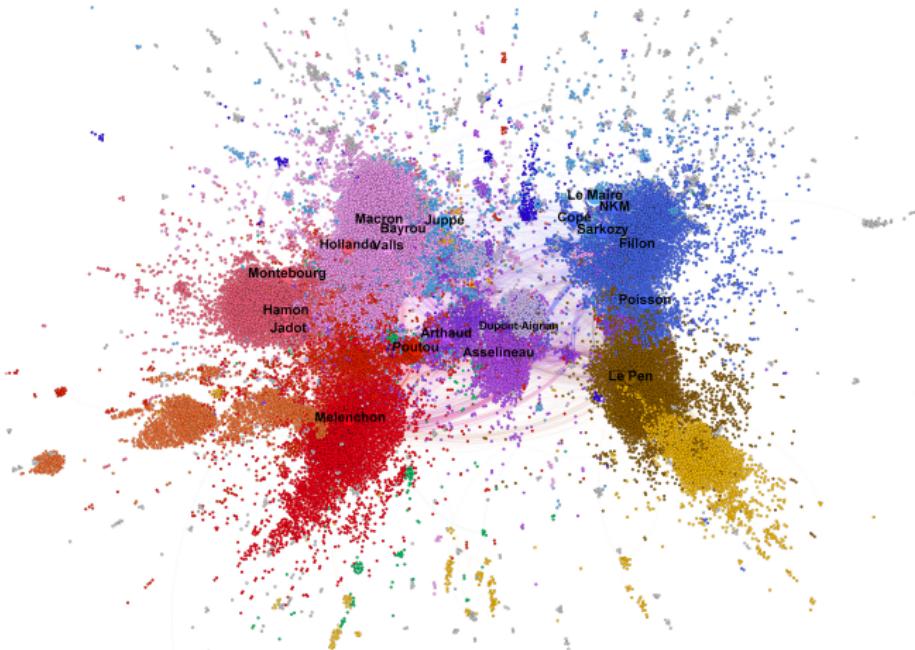


Figure: Avant le premier tour

IV) Quelques exemples d'usage des réseaux:

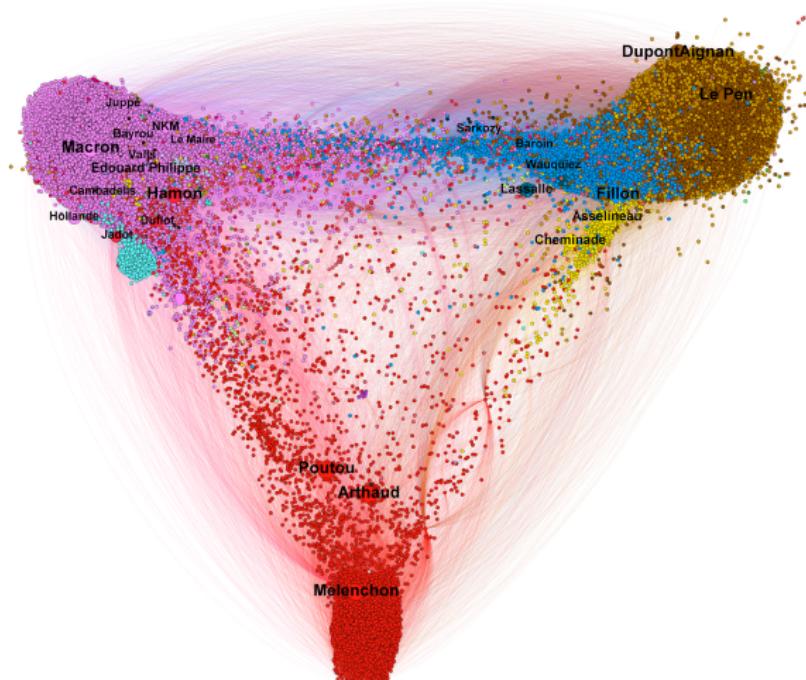


Figure: Avant le second tour

IV) Quelques exemples d'usage des réseaux:

"Opinions and political communities vary together. Political campaigns involve reversals, bombshells, and reconfigurations of political support. (...) The main result of our study is that the automatic analysis of retweet patterns is able to measure, sometimes before the media can make any announcement, the changes taking place inside political communities in response to these events and, in some cases, in response to anticipated events. These changes can be of large magnitude in a very short period of time."

Gaumont N., Panahi M., Chavalarias D., 2018, "Reconstruction of the socio-semantic dynamics of political activist Twitter networks—Method and application to the 2017 French presidential election", in *PLoS ONE*, 13(9).

IV) Quelques exemples d'usage des réseaux:

- Gaumont, Panahi, Chavalarias (2018): Étudient les réseaux socio-sémantiques politiques du twitter francophone pendant l'élection présidentielle française de 2017.
- Se focalisent sur les retweet de comptes politiques, et en infèrent une appartenance politique.
- Ils étudient la manière dont la structure du réseau de diffusion sur twitter évoluent durant toute la campagne. Ils traitent donc de l'aspect évolutif et longitudinal des réseaux.

Conclusion:

- Les réseaux sont une manière de représenter les structures relationnelles émergentes du monde social.
- Ce que vous pouvez dire de la structure du réseau dépend de vos choix méthodologiques: qui vous étudiez, quelles interactions, dans quel contexte?
- Il y a différentes manières d'étudier les réseaux. Plusieurs techniques de modélisation reposant sur diverses sous-structures existent.