

Detection automatique des structures communautaires dans les reseaux

Diop Abdoulaye
Encadré par Nicolas Verzelen

Département de Mathématiques
Université Montpellier

14 octobre 2024



STATISTIQUE
SCIENCE DES DONNÉES
UNIVERSITÉ DE MONTPELLIER

Introduction



Introduction



Interactions



**Contexte
scientifique**

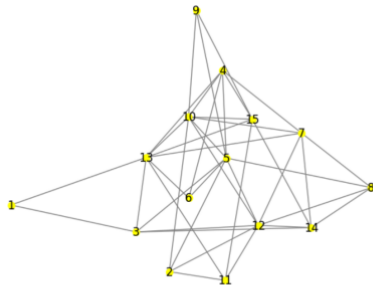


Problématique



**Méthodes
d'algorithme**

Soit un graphe $G = (V, E)$, où V est l'ensemble des sommets (nœuds) et E l'ensemble des arêtes (connexions) entre les sommets avec $V = \{1, 2, \dots, 15\}$ et $|E| = 40 = m$ (ensemble des nœuds).



(a) Graphe de 15 nœuds et 40 liens

	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14	N15
N1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
N2	0	0	0	0	1	0	0	0	0	1	1	1	0	0	0
N3	1	0	0	0	1	0	0	0	0	0	0	1	1	1	0
N4	0	0	0	0	1	1	1	0	0	1	0	0	1	0	1
N5	0	1	1	1	0	1	0	1	1	1	0	1	0	0	0
N6	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0
N7	0	0	0	1	0	0	0	1	0	1	0	1	1	1	0
N8	0	0	0	0	1	0	1	0	0	0	0	1	0	1	0
N9	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1
N10	0	1	0	1	1	0	1	0	1	0	0	1	1	0	1
N11	0	1	0	0	0	0	0	0	0	0	0	1	1	0	1
N12	0	1	1	0	1	0	1	1	0	1	1	0	0	1	0
N13	1	0	1	1	0	1	1	0	0	1	1	0	0	0	1
N14	0	0	1	0	0	0	1	1	0	0	0	1	0	0	1
N15	0	0	0	1	0	0	0	0	1	1	1	0	1	1	0

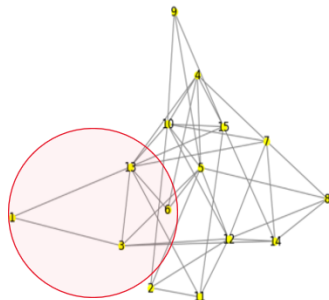
(b) Matrice d'adjacence du graphe

Figure 1: Représentation du graphe et de sa matrice d'adjacence.

Communauté dans un réseau

Soit un graphe $G = (V, E)$, où V est l'ensemble des sommets (nœuds) et E l'ensemble des arêtes (connexions) entre les sommets. Une communauté est un sous-ensemble de sommets $C \subseteq V$, tel que les connexions (arêtes) entre les sommets de C sont plus denses que les connexions entre C et $V \setminus C$ (le reste du graphe).

Exemple : $C = \{1, 3, 6, 13\}$



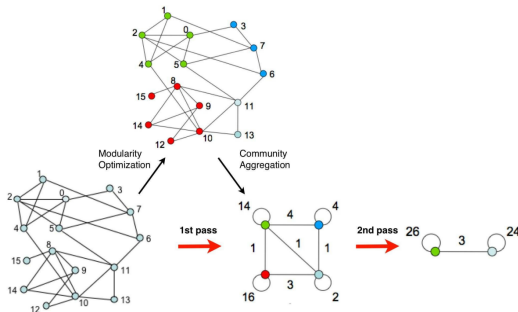
La modularité d'une partition

$$Q = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

où : A_{ij} est le poids de l'arête entre les nœuds i et j . Si i et j ne sont pas directement connectés, alors $A_{ij} = 0$. $k_i = \sum_j A_{ij}$ est la somme des poids des arêtes attachées au sommet i , c'est-à-dire le degré pondéré de i . c_i est la communauté à laquelle appartient le sommet i . $\delta(c_i, c_j)$ est une fonction indicatrice égale à 1 si $c_i = c_j$, c'est-à-dire si les nœuds i et j appartiennent à la même communauté, sinon elle vaut 0. $m = \frac{1}{2} \sum_{i,j} A_{ij}$ est le nombre total d'arêtes pondérées dans le réseau.

Méthode de Louvain

La méthode de Louvain est un algorithme de détection de communautés qui utilise une approche hiérarchique (agglomérative) pour maximiser la modularité d'un réseau. Il divise le réseau en groupes de nœuds, appelés communautés, de manière à ce que les nœuds à l'intérieur d'une communauté soient densément connectés entre eux, tandis que les connexions entre différentes communautés soient rares.



Étapes de la méthode

Étapes de l'algorithme

Phase 1 : Optimisation locale

Initialisation : Chaque nœud v_i est dans sa propre communauté $C_i = \{v_i\}$, $i = 1, 2, \dots, n$, où n est le nombre total de nœuds.

Modularité : Q

Déplacement : Pour chaque nœud v_i , calculez ΔQ si v_i est déplacé vers C_k :

$$\Delta Q = \left(\frac{P_{\text{in}} + k_{i,\text{in}}}{2m} - \frac{P_{\text{tot}} + k_i}{2m} \right)^2 - \left(\frac{P_{\text{in}}}{2m} - \left(\frac{P_{\text{tot}}}{2m} \right)^2 - \frac{k_i}{2m} \right)^2$$

avec P_{in} somme des poids des arêtes internes, $k_{i,\text{in}}$ degré de v_i vers C_k , et P_{tot} somme des poids de v_i vers tous les autres nœuds.

Répétez jusqu'à ce qu'il n'y ait plus d'amélioration de modularité.

Phase 2 : Agrégation

Nouveau réseau : Chaque communauté devient un nœud dans un nouveau réseau G' avec $A'_{kl} = \sum_{v_i \in C_k, v_j \in C_l} A_{ij}$, où A'_{kl} est le poids entre les communautés C_k et C_l .

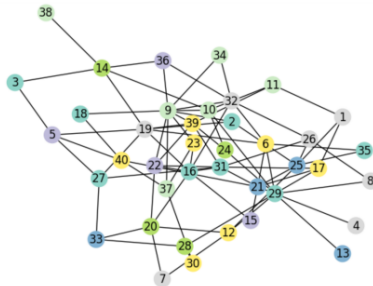
Répétez les phases 1 et 2 jusqu'à la convergence.



Applications de la méthode

Communauté 1: {32, 6, 11, 12, 19, 27, 28, 29}
Communauté 2: {16, 3, 20, 21, 30, 14, 15}
Communauté 3: {17, 36, 22, 39, 8}
Communauté 4: {0, 38, 23, 9, 10}
Communauté 5: {1, 34, 33, 5, 7, 24, 25, 31}
Communauté 6: {2, 35, 18, 37, 4, 26, 13}

(a) Communautés formées



(b) Clusters par couleurs pour un reseau de 40 noeuds et 80 liens

Comparaison de la méthode de Louvain avec d'autres méthodes

	Modularité (Louvain)	Modularité (CNM)
Réseau de petites tailles	0.1492	0.1468
Réseau de grandes tailles	0.4969	0.0001

	Karaté	Arxiv	Internet	Web nd.edu	Téléphone	Web uk-2005	Web WebBase 2001
Nœuds/liens	34/77	9k/24k	70k/351k	325k/1M	2.6M/6.3M	39M/783M	118M/1B
CNM	.38/0s	.772/3.6s	.692/799s	.927/5034s	-/-	-/-	-/-
PL	.42/0s	.757/3.3s	.729/575s	.895/6666s	-/-	-/-	-/-
WT	.42/0s	.761/0.7s	.667/62s	.898/248s	.56/464s	-/-	-/-
Notre algorithme	.42/0s	.813/0s	.781/1s	.935/3s	.769/134s	.979/738s	.984/152mn

Forces :

1. Algorithme intuitif et facile à mettre en œuvre.
2. Rapidité d'exécution avec une complexité généralement linéaire sur des données typiques et éparses.
3. Permet de détecter des communautés à différentes échelles.

Faiblesses :

1. Sensibilité à l'ordre des nœuds, pouvant influencer le temps de calcul.
2. Problème de résolution, limitant la détection de petites communautés.
3. Risque de se coincer dans des maxima locaux de la modularité.