

Théorie des réseaux

Chapitre 10 : La détection de communautés



Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger · 27 août 2026

La question

La modularité

Les algorithmes

La limite de résolution

Le cœur-périphérie

Ce qu'il faut retenir

La question

Une définition de travail

Un groupe de nœuds **plus densément reliés entre eux** qu'avec le reste du réseau.

Pourquoi la définition reste floue

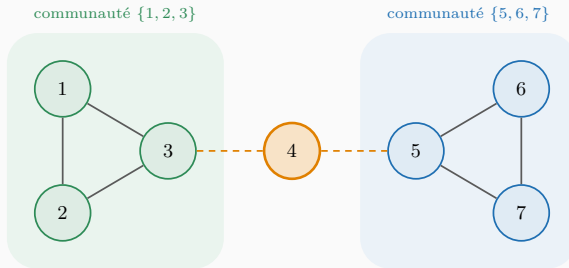
« Plus densément » par rapport à quoi ? Il n'existe pas de définition universelle, et chaque algorithme en implique une.

Deux méthodes différentes donnent deux partitions différentes du même réseau, sans qu'aucune soit fausse.

La partition évidente

$\{1, 2, 3\}$ et $\{5, 6, 7\}$

Reste le nœud 4 : il a un voisin de chaque côté. **Aucune méthode ne peut le placer sans arbitraire.**



$Q \approx 0,367$, quel que soit le côté où l'on range le nœud 4.
La modularité ne tranche pas : un pont n'appartient à aucune communauté.

La modularité

Newman et Girvan, 2004

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

$\delta(c_i, c_j)$ vaut 1 si i et j sont dans la même communauté.

Ce que la formule compare

A_{ij} est le lien **observé**. $k_i k_j / 2m$ est le lien **attendu** si les arêtes étaient recâblées au hasard en conservant les degrés.

Q est donc l'**excès de liens internes par rapport au hasard**, et non le nombre de liens internes.

L'ordre de grandeur

$Q \in [-0,5 ; 1]$. Une partition est jugée significative au-delà de $Q \approx 0,3$.

Partition en deux, le nœud 4 avec le premier groupe

Communautés $\{1, 2, 3, 4\}$ et $\{5, 6, 7\}$. Arêtes internes : 4 et 3; l'arête 4-5 est la seule coupée.

Somme des degrés : 9 et 7, donc

$$Q = \left(\frac{4}{8} - \left(\frac{9}{16} \right)^2 \right) + \left(\frac{3}{8} - \left(\frac{7}{16} \right)^2 \right) \approx 0,367$$

La formule pratique

$$Q = \sum_c \left[\frac{m_c}{m} - \left(\frac{d_c}{2m} \right)^2 \right]$$

m_c les arêtes internes à la communauté c , d_c la somme des degrés de ses membres.

Trois partitions comparées

Partition	Q
$\{1, 2, 3, 4\}$ et $\{5, 6, 7\}$	0,367
$\{1, 2, 3\}$ et $\{4, 5, 6, 7\}$	0,367
$\{1, 2, 3\}, \{4\}, \{5, 6, 7\}$	0,352

Ce que l'égalité signifie

Les deux premières partitions sont **strictement équivalentes** — la symétrie du réseau l'imposait.

La modularité ne tranche donc pas : elle indique que **la question est mal posée**. Un pont n'appartient à aucune communauté, et c'est une information, non une défaillance.

Les algorithmes

Blondel et al., 2008

Phase 1 — chaque nœud rejoint la communauté de son voisin qui augmente le plus Q , jusqu'à stabilité.

Phase 2 — on agrège chaque communauté en un super-nœud, et l'on recommence.

Pourquoi c'est la méthode standard

Complexité quasi linéaire : $O(n \log n)$. Elle traite des réseaux de dizaines de millions de nœuds.

Elle produit en outre une **hiérarchie** de partitions, du plus fin au plus grossier, ce qui est souvent l'information la plus utile.

Son défaut

Le résultat dépend de **l'ordre dans lequel les nœuds sont visités**.

Deux exécutions sur les mêmes données peuvent différer. **Il faut donc répéter et vérifier la stabilité** — un rapport qui donne une seule exécution n'a rien montré.

2002 - par les ponts

On retire itérativement l'arête de plus forte **intermédiarité d'arête**, en recalculant à chaque fois.

Le graphe se fragmente; on conserve la partition de meilleure modularité.

Sur notre réseau

Les arêtes 3-4 et 4-5 sont celles de plus forte intermédiarité : ce sont les ponts.

Les retirer isole $\{1, 2, 3\}$, $\{4\}$ et $\{5, 6, 7\}$. L'algorithme fait du nœud 4 une communauté à lui seul — ce qui est une réponse honnête.

Son coût

$O(n m^2)$: inutilisable au-delà de quelques milliers de nœuds.

On l'emploie sur de petits réseaux, pour son interprétabilité : **il dit quelles arêtes séparent**, pas seulement quels groupes existent.

La limite de résolution

Fortunato et Barthélemy, 2007

La modularité **ne détecte pas** les communautés plus petites que $\sqrt{2m}$, quelle que soit leur densité interne.

Pourquoi

Le terme attendu $k_i k_j / 2m$ dépend de la taille du réseau entier.

Sur un grand réseau, deux petites cliques parfaitement séparées et reliées par une seule arête donnent une modularité **plus élevée fusionnées que séparées**.

La conséquence pratique

Maximiser Q sur un réseau d'un million de nœuds fusionne systématiquement les petits groupes.

Le remède : introduire un paramètre de résolution γ dans le terme attendu, et examiner la partition à plusieurs échelles plutôt qu'à une seule.

Le cœur-périphérie

Ce que ce n'est pas

Une structure en communautés suppose des groupes **entre lesquels** les liens sont rares.

Une structure **cœur-périphérie** a un noyau dense et une périphérie qui se lie au cœur mais **pas à elle-même**.

Où on la trouve

Le marché interbancaire : quelques dizaines de grandes banques échangent entre elles, des centaines de petites passent par elles.

Appliquer un algorithme de communautés à un réseau cœur-périphérie produit une partition qui n'a **aucun sens**. Il faut tester la structure avant de choisir la méthode.

Ce qu'il faut retenir

Six points

1. Une communauté est plus dense en interne qu'avec l'extérieur — définition **non universelle**.
2. La **modularité** compare l'observé au hasard à degrés conservés, pas au hasard uniforme.
3. Sur notre réseau, la partition à deux groupes donne $Q \approx 0,367$.
4. **Louvain** est rapide et hiérarchique, mais dépend de l'ordre de visite.
5. **Girvan-Newman** retire les ponts : interprétable, mais en $O(n m^2)$.
6. La **limite de résolution** interdit de détecter les groupes plus petits que $\sqrt{2m}$.

La suite

Les dix premiers chapitres ont décrit des réseaux. Le chapitre 11 y met quelque chose en mouvement : une maladie, une rumeur, une faillite.