

Théorie des réseaux

Chapitre 3 : La matrice d'adjacence et l'algèbre des graphes



La matrice d'adjacence

Les puissances comptent les chemins

Les autres matrices

Le spectre

Ce qu'il faut retenir

La matrice d'adjacence

A

$$A_{ij} = \begin{cases} 1 & \text{si } ij \in E \\ 0 & \text{sinon} \end{cases}$$

Pour un graphe non orienté, **A** est **symétrique** et sa diagonale est nulle.

Notre réseau

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

Trois lectures

Le **degré** est la somme d'une ligne : $k_i = \sum_j A_{ij}$.

Le **nombre d'arêtes** vaut $m = \frac{1}{2} \sum_{ij} A_{ij}$.

Le vecteur des degrés s'écrit $\mathbf{k} = \mathbf{A}\mathbf{1}$.

Une matrice creuse

Sur 49 cases, 16 valent 1. Sur un réseau de 10^5 nœuds et 10^6 arêtes, la proportion tombe à 2×10^{-4} .

On ne stocke jamais la matrice d'un grand réseau : on stocke la liste des arêtes.

Les puissances comptent les
chemins

Le théorème

$$(\mathbf{A}^r)_{ij} = \text{nombre de parcours de longueur exactement } r \text{ de } i \text{ vers } j$$

Un **parcours** peut repasser par un sommet — contrairement à un chemin.

La démonstration, par récurrence

$$(\mathbf{A}^2)_{ij} = \sum_h A_{ih} A_{hj}$$

Chaque terme vaut 1 si h est voisin de i et de j . La somme compte donc les intermédiaires possibles — c'est-à-dire les parcours de longueur 2.

Trois lectures des puissances

$(\mathbf{A}^2)_{ii} = k_i$: les parcours de longueur 2 qui reviennent sont les aller-retours vers un voisin.

$(\mathbf{A}^3)_{ii} = 2 \times$ le nombre de **triangles** contenant i .

$\text{trace}(\mathbf{A}^3) = 6 \times$ le nombre total de triangles.

Sur notre réseau

Deux triangles : $\{1, 2, 3\}$ et $\{5, 6, 7\}$.

$$\text{trace}(\mathbf{A}^3) = 6 \times 2 = 12$$

L'intérêt pratique

Compter les triangles d'un réseau d'un million de nœuds par énumération est hors de portée.

Le compter par une trace de produit matriciel est une opération d'algèbre linéaire creuse — quelques secondes.

Les autres matrices

La définition

$$\mathbf{L} = \mathbf{D} - \mathbf{A}, \quad \mathbf{D} = \text{diag}(k_1, \dots, k_n)$$

$\mathbf{L}$ est symétrique, semi-définie positive, et la somme de chaque ligne est nulle.

Sa propriété fondamentale

$$\mathbf{L}\mathbf{1} = \mathbf{0}$$

donc 0 est toujours valeur propre. Sa multiplicité est le nombre de composantes connexes.

Ce que cela permet

Compter les composantes d'un réseau revient à compter les zéros du spectre de $\mathbf{L}$.

La **deuxième** plus petite valeur propre — la connectivité algébrique de Fiedler — mesure à quel point le graphe est **difficile à couper**. C'est l'outil du chapitre 10.

La marche aléatoire

$$P_{ij} = \frac{A_{ij}}{k_i}$$

P_{ij} est la probabilité qu'un marcheur en i passe en j . Chaque ligne somme à 1.

Sa loi stationnaire

Sur un graphe connexe non biparti, la marche converge vers

$$\pi_i = \frac{k_i}{2m}$$

La fraction du temps passée en un nœud est proportionnelle à son degré. C'est l'idée dont PageRank est une variante — et c'est le chapitre 5.

Le spectre

Lambda 1

La plus grande valeur propre de $\mathbf{A}$. Pour un graphe connexe, elle est **simple** et son vecteur propre est **strictement positif** — théorème de Perron-Frobenius.

Un encadrement

$$\bar{k} \leq \lambda_1 \leq k_{\max}$$

Sur notre réseau

$$\bar{k} = 2,29 \leq \lambda_1 \approx 2,343 \leq 3 = k_{\max}$$

Le polynôme caractéristique se réduit, par symétrie du réseau, à $\lambda^3 - \lambda^2 - 4\lambda + 2 = 0$.

Trois rôles

Le seuil épidémique — une contagion se propage si le taux de transmission dépasse $1/\lambda_1$ (chapitre 11).

La centralité de vecteur propre — le vecteur propre associé la définit (chapitre 5).

La stabilité d'un jeu en réseau — l'équilibre existe et est unique si λ_1 est assez petit (chapitre 12).

Une leçon de méthode

Trois questions économiques sans rapport apparent — épidémie, pouvoir, équilibre — dépendent du même nombre.

C'est l'argument le plus fort en faveur de l'approche par les réseaux : elle unifie ce que l'intuition sépare.

Ce qu'il faut retenir

Six points

1. $\mathbf{A}$ code le graphe; $\mathbf{k} = \mathbf{A}\mathbf{1}$, $m = \frac{1}{2} \sum A_{ij}$.
2. $(\mathbf{A}^r)_{ij}$ compte les **parcours de longueur r** .
3. $\text{trace}(\mathbf{A}^3) = 6 \times$ le nombre de triangles – ici 12.
4. Le **laplacien** $\mathbf{L} = \mathbf{D} - \mathbf{A}$ compte les composantes par ses zéros.
5. La marche aléatoire a pour loi stationnaire $\pi_i = k_i/2m$.
6. $\lambda_1 \approx 2,343$ gouvernera le seuil épidémique, la centralité spectrale et l'équilibre des jeux.

La suite

Le chapitre 4 pose la question centrale du cours : **qui est important dans un réseau?** Il y répond de trois façons différentes, qui ne donnent pas le même classement.