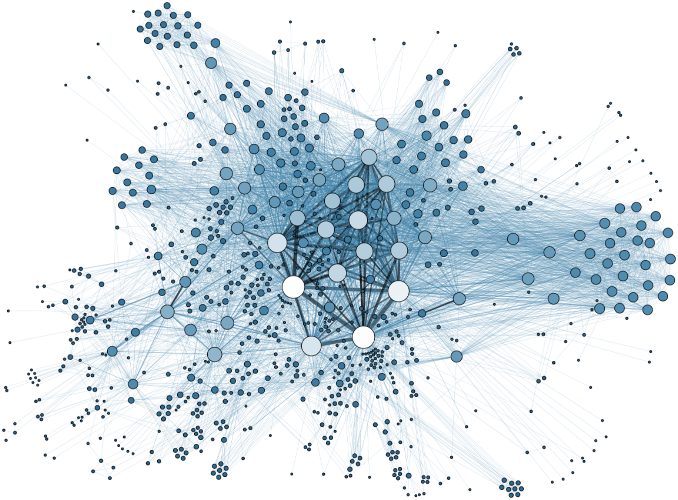
Tous connectés sur les réseaux sociaux.
Quels outils mathématiques pour les analyser ?

Tabea Rebafka

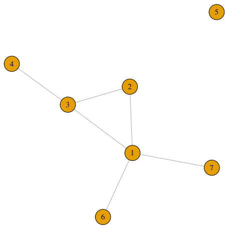
25 Novembre 2023

Mathematic Park

Outline



Définition d'un graphe



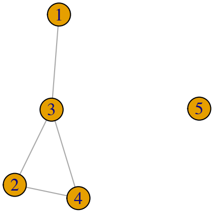
Graphe

Un **graphe** $G = (V, E)$ est composé

- d'un ensemble $V = \{1, \ldots, n\}$ de **nœuds** ou **sommets** (*vertices*) qui représentent les individus ou entités qui interagissent entre eux, et
- d'un ensemble E d'**arêtes** (*edges*) qui indiquent la présence d'une interaction ou connexion entre deux nœuds : $\{i, j\} \in E$ s'il y a une arête entre les nœuds i et j dans G.

Représentations équivalentes d'un graphe

Ensemble de nœuds
$V = \{1, \ldots, 5\}$
Ensemble d'arêtes
$E = \{(1,3), (2,3), (2,4), (3,4)\}$



Matrice d'adjacence

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

Outline

Génération aléatoire de graphes

Modèle de Erdös-Rényi ou graphe $G(n, p)$

- Choisir un nombre de noeuds n et une probabilité de connexion p
- Les arêtes du graphes sont tirées au hasard et indépendamment avec probabilité p :
  Les entrées $A_{i,j}$ de la matrice d'adjacence sont des variables aléatoires indépendantes de loi de Bernoulli de paramètre p :

$$A_{i,j} \sim \text{Bernoulli}(p), \quad \text{pour tout } i < j$$

et on pose $A_{i,j} = A_{j,i}$ pour tout $i > j$.

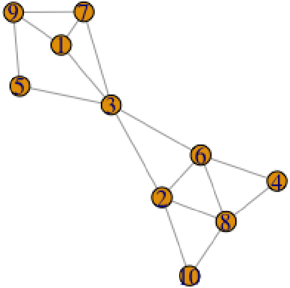
Outline

Détection de communautés I

Objectif

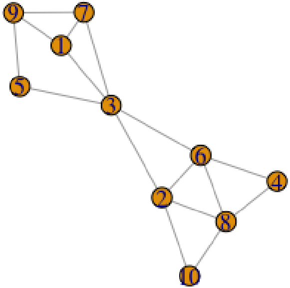
- Partitionner les nœuds du graphe en un nombre fini de groupes
- Les nœuds d'un groupe doivent être plus connectés entre eux qu'avec les nœuds du reste du graphe.

Détection de communautés II



0	0	1	0	0	0	1	0	1	0
0	0	1	0	0	1	0	1	0	1
1	1	0	0	1	1	1	0	0	0
0	0	0	0	0	1	0	1	0	0
0	0	1	0	0	0	0	0	1	0
0	1	1	1	0	0	0	1	0	0
1	0	1	0	0	0	0	0	1	0
0	1	0	1	0	1	0	0	0	1
1	0	0	0	1	0	1	0	0	0
0	1	0	0	0	0	0	1	0	0

Détection de communautés III



0	0	1	0	0	0	1	0	1	0
0	0	1	0	0	1	0	1	0	1
1	1	0	0	1	1	1	0	0	0
0	0	0	0	0	1	0	1	0	0
0	0	1	0	0	0	0	0	1	0
0	1	1	1	0	0	0	1	0	0
1	0	1	0	0	0	0	0	1	0
0	1	0	1	0	1	0	0	0	1
1	0	0	0	1	0	1	0	0	0
0	1	0	0	0	0	0	1	0	0

0	1	0	1	1	0	0	0	0	0
1	0	1	1	0	1	0	1	0	0
0	1	0	0	1	0	0	0	0	0
1	1	0	0	1	0	0	0	0	0
1	0	1	1	0	0	0	0	0	0
0	1	0	0	0	0	0	1	1	1
0	0	0	0	0	0	0	1	1	0
0	1	0	0	0	1	1	0	1	0
0	0	0	0	0	1	1	1	0	1
0	0	0	0	0	1	0	0	1	0

Détection de communautés IV

Heuristique du spectral clustering

- Si le graphe est composé de communautés, alors il existe une permutation des lignes et des colonnes de la matrice d'adjacence pour laquelle cette matrice est (presque) **diagonale par bloc**.
- On peut diagonaliser la matrice d'adjacence pour trouver cette permutation.
- Approche plus robuste : diagonaliser une transformation de la matrice d'adjacence : la **matrice laplacienne** du graphe

Matrices laplaciennes de graphe I

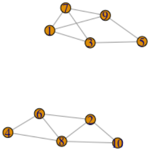
- Soit A la matrice d'adjacence d'un graphe non dirigé
- Notons D la **matrice des degrés** définie comme la matrice diagonale de taille n dont la diagonale vaut $(d_1, \ldots, d_n)$, où $d_i = \sum_j A_{ij}$ est le degré du noeud i

Laplacien

On définit la **matrice laplacienne** L d'un graphe par

$$L = D - A$$

Matrices laplaciennes de graphe II



$$C_1 = \{1,3,5,7,9\} \qquad \mathbb{1}_{C_1} = (1,0,1,0,1,0,1,0,1,0)^T$$
$$C_2 = \{2,4,6,8,10\} \qquad \mathbb{1}_{C_2} = (0,1,0,1,0,1,0,1,0,1)^T$$

Théorème

- Notons k le nombre de composantes connexes, $C_1, \dots, C_k$ les composantes connexes et $\mathbb{1}_{C_1}, \dots, \mathbb{1}_{C_k} \in \mathbb{R}^n$ les vecteurs indicatrices des composantes.
- Alors, toutes les valeurs propres de L sont positives ou nulles et $k =$ la multiplicité de la valeur propre 0
- En plus, $\mathbb{1}_{C_1}, \dots, \mathbb{1}_{C_k}$ sont des vecteurs propres associés à la valeur propre 0. Plus précisément, si v est vecteur propre associé à 0 (càd $Lv = 0$), alors v est de la forme :
  $$v = \sum_{\ell=1}^{k} a_\ell \mathbb{1}_{C_\ell} \quad \text{pour des constantes } a_1, \dots, a_k.$$