La théorie des graphes

Présentation et exercices

Martial Foegel

Laboratoire de Linguistique Formelle

Théorie des graphes et réseaux complexes

Exemple d'un phénomène complexe

Analyse d’un réseau complexe

1. **Mesure:** observe un phénomène dans la nature

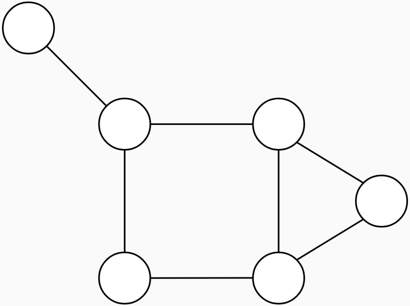
Analyse d’un réseau complexe

1. **Mesure:** observe un phénomène dans la nature
2. **Modèle:** essaye d’expliquer ses observations
    2.1 choisissant le bon niveau de grainage pour les sommets et les arrêtes
    2.2 réduit le problème à sa forme la plus simple
    2.3 formule le problème en terme mathématique

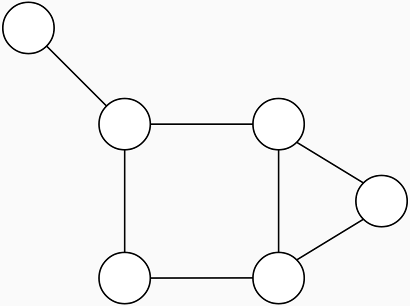
Analyse d'un réseau complexe

1. **Mesure:** observe un phénomène dans la nature
2. **Modèle:** essaye d'expliquer ses observations
   2.1 choisissant le bon niveau de grainage pour les sommets et les arrêtes
   2.2 réduit le problème à sa forme la plus simple
   2.3 formule le problème en terme mathématique
3. **Validation:** vérifie si les observations sont valides ou reproductibles

Exemple d’un graphe



Exemple d’un graphe



Vocabulary / vocabulaire:

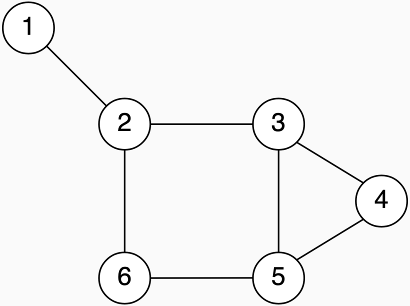
- Graph or network / graphe ou réseau
- Vertices, nodes or points / sommets, noeuds ou points
- Edges, links or lines / arrêtes, liens ou lignes

Mesures pour caractériser un graphe

Représentation d'un graphe

$G = (V, E)$,

avec V un set de N sommets (vertices) $V = \{v_1, \cdots v_n\}$ et E un set d'arrêtes (edges) v_i, v_j avec $v_i, v_j \in V$ et $v_i \neq v_j$.



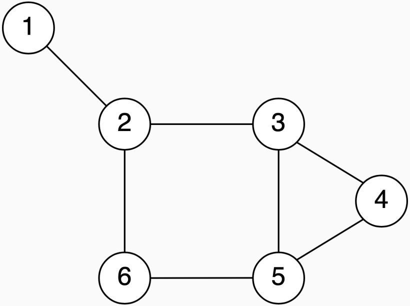
Représentation d'un graphe

$G = (V, E)$,

avec V un set de N sommets (vertices) $V = \{v_1, \cdots v_n\}$ et E un set d'arrêtes (edges) v_i, v_j avec $v_i, v_j \in V$ et $v_i \neq v_j$.

Matrice d'adjacence $A = a_{ij}$:

	1	2	3	4	5	6
1	0	1	0	0	0	0
2	1	0	1	0	0	1
3	0	1	0	1	1	0
4	0	0	1	0	1	0
5	0	0	1	1	0	1
6	0	1	0	0	1	0



Distribution des degrés

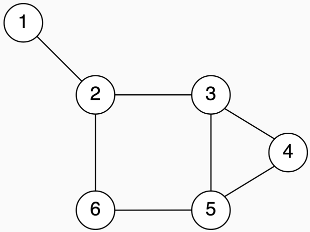
Le degré d'un sommet et le nombre de connections de ce sommet

$$k_i = a_{i1} + a_{i2} + ... + +a_{in} = \sum_{j}^{N} a_{ij},$$

où k_i est le dégré du sommet i dans la matrice d'adjacence A.

De plus, $m = \frac{\sum_i k_i}{2}$ où m est le nombre de sommets.

	1	2	3	4	5	6
1	0	1	0	0	0	0
2	1	0	1	0	0	1
3	0	1	0	1	1	0
4	0	0	1	0	1	0
5	0	0	1	1	0	1
6	0	1	0	0	1	0



Distribution des degrés

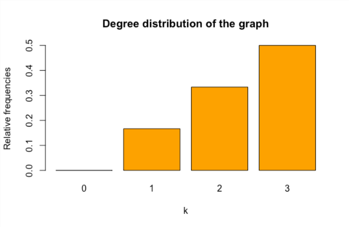
Le degré d'un sommet et le nombre de connections de ce sommet

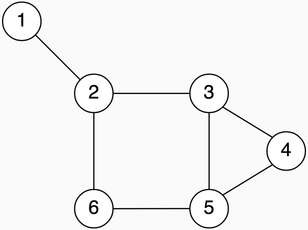
$$k_i = a_{i1} + a_{i2} + \ldots + +a_{in} = \sum_{j}^{N} a_{ij},$$

où k_i est le dégré du sommet i dans la matrice d'adjacence A.

De plus, $m = \frac{\sum_i k_i}{2}$ où m est le nombre de sommets.

	1	2	3	4	5	6
1	0	1	0	0	0	0
2	1	0	1	0	0	1
3	0	1	0	1	1	0
4	0	0	1	0	1	0
5	0	0	1	1	0	1
6	0	1	0	0	1	0





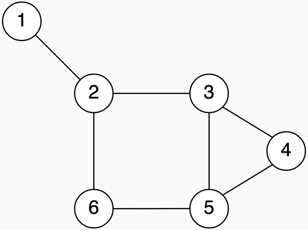
Moyenne du coefficient de clustering local

Quel portion des voisins d'un sommet sont connectés entre eux ?
Coefficient de clustering local:

$$C_u = \frac{2e_u}{k_u(k_u - 1)},$$

où e_u est le nombre de lien entre voisins et k_u est le nombre de voisin du sommet u, avec $C_u = 0$ si $k_u \in \{0, 1\}$.

	1	2	3	4	5	6
1	0	1	0	0	0	0
2	1	0	1	0	0	1
3	0	1	0	1	1	0
4	0	0	1	0	1	0
5	0	0	1	1	0	1
6	0	1	0	0	1	0



Moyenne du coefficient de clustering local

Quel portion des voisins d'un sommet sont connectés entre eux ?
Coefficient de clustering local:

$$C_u = \frac{2e_u}{k_u(k_u-1)},$$

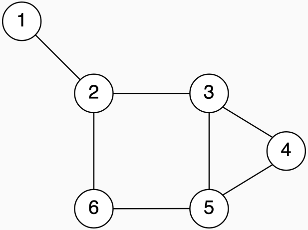
où e_u est le nombre de lien entre voisins et k_u est le nombre de voisin du sommet u, avec $C_u = 0$ si $k_u \in \{0, 1\}$.

	1	2	3	4	5	6
1	0	1	0	0	0	0
2	1	0	1	0	0	1
3	0	1	0	1	1	0
4	0	0	1	0	1	0
5	0	0	1	1	0	1
6	0	1	0	0	1	0

Moyenne du coefficient de clustering local:

$$C = \frac{1}{N}\sum_u C_u$$

$$Here, C = \frac{1}{6}\left(0+0+\frac{1}{3}+1+\frac{1}{3}+0\right) = \frac{5}{18}$$



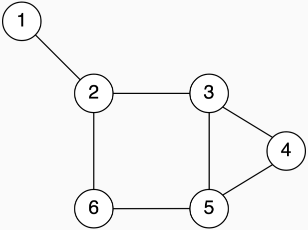
Distance moyenne minimum entre paires de sommets

Moyenne sur toutes les paires de sommets, du nombre de sommets se trouvant sur le chemin le plus court les séparants.

$$\overline{d} = \frac{2}{N(N-1)} \sum_{i<j} d_{ij},$$

où d_{ij} est la distance la plus courte entre i et j.

	1	2	3	4	5	6
1	0	1	0	0	0	0
2	1	0	1	0	0	1
3	0	1	0	1	1	0
4	0	0	1	0	1	0
5	0	0	1	1	0	1
6	0	1	0	0	1	0



Distance moyenne minimum entre paires de sommets

Moyenne sur toutes les paires de sommets, du nombre de sommets se trouvant sur le chemin le plus court les séparants.

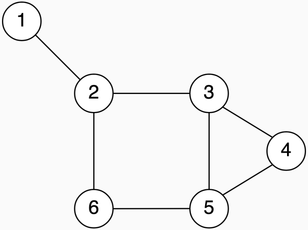
$$\overline{d} = \frac{2}{N(N-1)} \sum_{i<j} d_{ij},$$

où d_{ij} est la distance la plus courte entre i et j.

	1	2	3	4	5	6
1	0	1	0	0	0	0
2	1	0	1	0	0	1
3	0	1	0	1	1	0
4	0	0	1	0	1	0
5	0	0	1	1	0	1
6	0	1	0	0	1	0

La matrices de distance :

	2	3	4	5	6
1	1	2	3	3	2
2		1	2	2	1
3			1	1	2
4				1	2
5					1



Nous finissons avec $\overline{d} = \frac{5}{3}$

Passons à la pratique avec le package *igraph* sur R