Correction TME 5

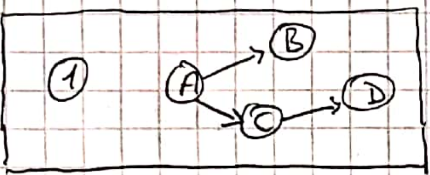
3) Degree distribution

a) Soit $G=(V,E)$ un graphe, avec $\begin{cases} V \text{ l'ensemble de ses noeuds} \\ E \text{ l'ensemble de ses arêtes} \end{cases}$

On dit que : $\begin{cases} G \text{ est dirigé si les arcs ont une direction, comme en figure ①} \\ G \text{ est non-dirigé sinon.} \end{cases}$

b) Degré d'un noeud

Définition : Le degré d'un noeud v ($\deg(v)$) est le nombre de connections de ce noeud à d'autres noeuds (nombre d'arcs reliés à v).



$\begin{cases} \deg(A) = 2 \\ \deg(B) = 1 \end{cases}$

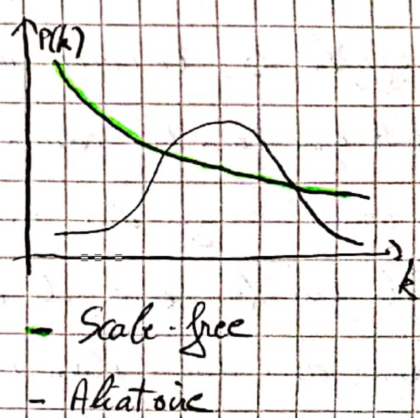
Distribution de degrés $P(k)$: distribution de probabilité des degrés dans un graphe.

↳ Réseau scale-free : cette distribution suit une loi de puissance, c'est-à-dire qu'il existe $\gamma \in \mathbb{N}$ tel que $P(k) \sim k^{-\gamma}$ (typiquement $2 < \gamma < 3$).

Propriétés d'un tel réseau : 1) Les hubs sont souvent entourés de noeuds de bas degré
2) Pas mal de noeuds à degré très élevé (voir figure)

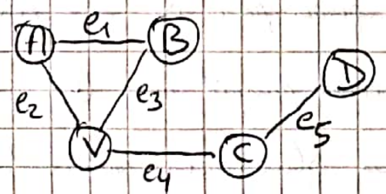
↳ Réseau aléatoire : $P(k) \sim \binom{n-1}{k} p^k (1-p)^{n-1-k}$ où n est le nombre de noeuds de G (loi binomiale)

Intérêt de savoir si un réseau est scale-free : Robustesse du réseau. Si l'on enlève un noeud/arc au hasard, la probabilité de perdre la connectivité est faible comparée à celle d'un réseau aléatoire.



4) Clustering coefficient

Exemple :



Quel est le clustering coef. de v ?

a) Voisins de v : $N(v) = \{A, B, C\}$

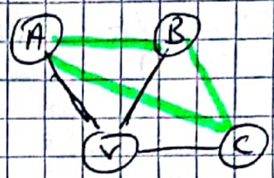
~~b) E(N(v)) = {(u1, u2) ∈ E | (u1, u2) ∈~~

b) $E(N(v)) = \{(u_1, u_2) \in E \mid u_1, u_2 \in \{A,B,C\}\}$

donc ~~E~~ : $\begin{cases} e_1 \in E(N(v)) & \text{car A et B appartient à } \{A,B,C\} \\ e_2 \notin E(N(v)) & \text{car } v \notin \{A,B,C\} \\ e_3 \notin E(N(v)) & \end{cases}$

Finalement : $E(N(v)) = \{e_1\}$, donc $|E(N(v))| = 1$

Or $\deg(v) = 3$, donc $cc(A) = \frac{2\times 1}{3\times 2} = \frac{1}{3}$



Alternativement : Combien de triangles passent par v ? → 1

Combien de triangles pourraient passer par v ? 3

donc $cc(v) = \frac{1}{3}$

⚠ Pour un graphe dirigé : $cc(v) = \frac{|E(N(v))|}{\deg(v)(\deg(v)-1)}$

Pour le réseau G :

$cc(A) = 0$ $\quad cc(B) = \frac{4\times 2}{10\times 9}$, $cc(C) = \frac{2\times 2}{10\times 9}$, $cc(B) = \frac{2\times 1}{2\times 1}$, $cc(E) = \frac{2\times 6}{10\times 9}$

Intérêt : Mettre en évidence un réseau scale-free.

Connectivité : Nombre minimal d'arcs/noeuds à enlever pour séparer le graphe en deux sous-graphes isolés. L'intérêt de les ~~tro~~ identifier est de trouver les points critiques du réseau, les endroits les plus vulnérables aux attaques.

Connectivité faible ⇔ résilience faible.

⇒ Comment les trouver ? → betweenness centrality

Betweenness centrality permet de trouver les noeuds centraux au réseau. Dans le graphe G, le noeud intermédiaire V liant ~~à~~ A et B a ~~un noeud~~ une BC élevée, parce que tous les noeuds autour de A doivent passer par V pour passer vers le reste du graphe. Il y a donc beaucoup de plus court chemins passant par V. Identifier des noeuds tels que V permet de mettre en évidence des structures communautaires (groupe de noeuds avec densité élevée) telles que celles autour de A, B, C ou E dans le graphe G.