AAGB – TME 5

Réseaux

1 Introduction

La théorie des réseaux consiste en l'étude des propriétés statistiques qui caractérisent la structure et le comportement des systèmes définis par des réseaux. On propose d'étudier quelques mesures classiques pour les réseaux : distribution de degré, coefficient de clustering et betweenness centrality. Pour tester vos fonctions, vous pourrez par exemple utiliser les deux fichiers (reseau1, reseau2) qui listent l'ensemble des arêtes présentes dans un réseau. On peut par exemple parser ces fichiers et créer un dictionnaire avec pour clés les noeuds du réseau, et pour valeurs la liste des noeuds reliés au noeud en question.

2 Degree distribution

1. Définissez la distribution de degrés pour un graphe.
2. Quel est l'intérêt d'étudier la distribution des degrés dans un graphe ?
3. Tracez la distribution des degrés de ce réseau avec échelle log-log.
4. Qu'est ce qu'un réseau scale-free ? Parmi les deux réseaux étudiés, lequel pourrait être scale-free ? Lequel aléatoire ? Pourquoi ?

3 Clustering coefficient

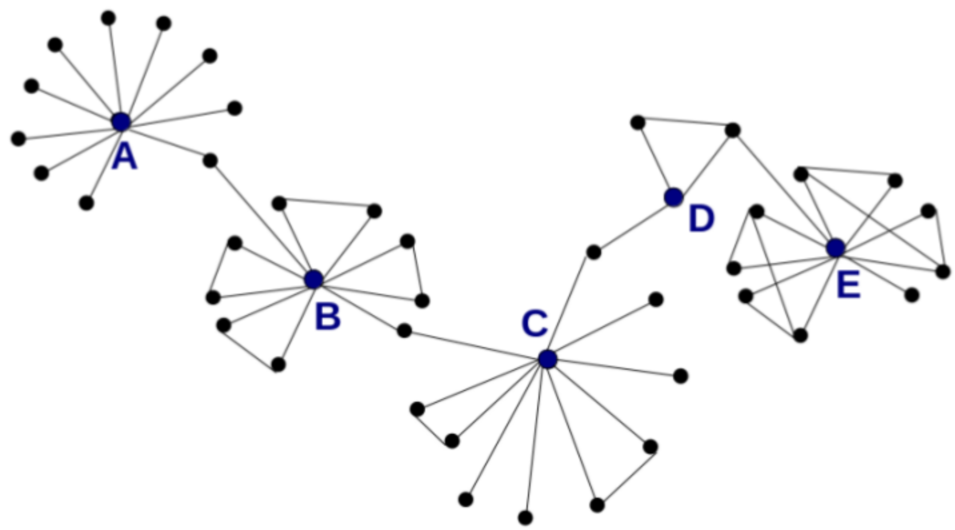
Le *coefficient de clustering* donne une mesure *locale* de l'interconnectivité des voisins d'un sommet. Pour un réseau scale-free, la distribution log-log des moyennes des coefficients de clustering, $C(k)$, en fonction du degré des sommets, k, suit une loi de puissance du type $C(k) \propto k^{-\gamma}$. Cela indique que certains nœuds du réseau se regroupent en clusters (contrairement aux réseaux aléatoires).

Définition 1 (Clustering Coefficient) *Soit $G = (V, E)$ un graphe et $v \in V$ un sommet. Soit $N(v)$ l'ensemble des voisins de v et $E(N(v)) = \{(u_1, u_2) \in E \mid u_1, u_2 \in N(v)\}$. Le coefficient de clustering $cc(v)$ de v est défini par*

$$cc(v) = \frac{2 \mid E(N(v)) \mid}{d^\circ(v)(d^\circ(v) - 1)}$$

En d'autres termes, $cc(v)$ est le ratio entre le nombre de 'triangles' passant par le sommet v ($E(N(v))$) et le nombre de 'triangles' qui pourraient passer par v ($d^\circ(v)(d^\circ(v) - 1)/2$).

Considérons le réseau G suivant :



1. Calculez $cc\left(A\right), cc\left(B\right), cc\left(C\right), cc\left(D\right)\ et\ cc\left(E\right)$.
2. Ecrivez une fonction permettant de calculer la moyenne des coefficient de clustering. Que peut on en déduire pour les réseaux étudiés ?

4 Betwenness centrality

La betweenness centrality (ou centralité intermédiaire) est une propriété qui tente de capturer l'importance d'un nœud (ou d'un lien) pour la propagation d'un signal/message, autrement dit son importance en tant qu'intermédiaire dans la structure du réseau. Elle repose essentiellement sur la notion de plus court chemin (ie. le chemin comportant le moins d'arcs possible).

Par convention, on notera σ_{st} le nombre de plus courts chemins entre les nœuds s et t d'un graphe G et $\sigma_{st}\left(v\right)$ le nombre de plus courts chemins entre les nœuds s et t passant par v (on suppose ici que $s \neq t \neq v$). La betweenness centrality d'un nœud v est alors définie par la formule suivante :

$$BC\left(v\right) = \sum_{s\neq t\neq v} \frac{\sigma_{st}\left(v\right)}{\sigma_{st}}$$

Un cas particulier simple pour le calcul de $BC\left(v\right)$ se présente lorsqu'il existe exactement 1 plus court chemin pour toutes les paires de sommets (s,t) de G. Dans ce cas, $BC\left(v\right)$vaut exactement $BC\left(v\right) = \sum_{s\neq t\neq v} \sigma_{st}\left(v\right)$, c'est-à-dire le nombre de plus courts chemins dans G passant par v.

1. Quel type de noeuds permet de mettre en évidence la betweenness centrality ? Un réseau comportant quelques noeuds avec une forte centralité est-il robuste ? Comment utiliser cette mesure pour détecter des communautés ?
2. Ecrire une fonction permettant de calculer la betweenness centrality.